

QY 1 ATGRLLSGQPVCRGGTQPCYKVIYFHDTSRRLNFEEAKEACRRDGGQLVSI ESEDEQKL 60
|||||
Db 22 ATGRLLSGQPVCRGGTQPCYKVIYFHDTSRRLNFEEAKEACRRDGGQLVSI ESEDEQKL 81
|||||
QY 61 IEKFIENLLPSDGFWIGLRRREKQSNSTACQDLYAWTDGSI SQFRNMYVDEPSCGSEV 120
|||||
Db 82 IEKFIENLLPSDGFWIGLRRREKQSNSTACQDLYAWTDGSI SQFRNMYVDEPSCGSEV 141
|||||
QY 121 CVVMYHOPSAPAGIGGYPYMFQWDDRCNMKNNFICKYSDEKPAVPSREAEGEETELTPV 180
|||||
Db 142 CVVMYHOPSAPAGIGGYPYMFQWDDRCNMKNNFICKYSDEKPAVPSREAEGEETELTPV 201
|||||
QY 181 LPEETQEEDAKKTFKESREAAALNLAY 206
|||||
Db 202 LPEETQEEDAKKTFKESREAAALNLAY 227
|||||
RESULT 3
ABG66680
ID ABG66680 standard; Protein; 374 AA.
XX
AC ABG66680;
XX
DT 30-AUG-2002 (first entry)
XX
DE Human novel polypeptide #15.
XX
KW Human; inflammatory condition; shock; sepsis; immune response;
KW cancer; wound healing; central nervous system disease; haematopoiesis;
KW peripheral nervous system disease; amyotrophic lateral sclerosis; tendon;
KW myeloid cell disorder; lymphoid cell disorder; platelet disorder; bone;
KW cartilage; ligament; nerve tissue; ulcer; osteoporosis; osteoarthritis;
KW bone degenerative disorder; periodontal disease; reperfusion injury;
KW lung fibrosis; liver fibrosis; autoimmune disorder; bacterial infection;
KW allergic condition; thrombolysis; thrombosis; coagulation disorder;
KW fungal infection.
XX
OS Homo sapiens.
XX
PN WO200244340-A2.
XX
PD 06-JUN-2002.
XX
PF 30-NOV-2001; 2001WO-US47004.
XX
PR 30-NOV-2000; 2000US-0028952.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Tang YT, Goodrich RW, Liu C, Zhou P, Asundi V, Wang J, Wang D;
PI Yamazaki V, Ujwal ML, Drmanac RT;
XX
DR WPI; 2002-508509/54.
DR N-PSDB; ABK94904.
XX
PT Novel nucleic acids and polypeptides for diagnosis, treatment of
PT inflammatory, autoimmune, nervous system, myeloid or lymphoid cell
PT disorders, cancer and promoting wound healing -
XX
PS Claim 10; Page 579-580; 672pp; English.
XX
CC The invention relates to human novel polynucleotides and associated
CC polypeptides. The polynucleotides and polypeptides are useful for
CC treating inflammatory conditions such as arthritis, nephritis, Crohn's
CC disease, ischaemia-reperfusion injury, shock, sepsis, immune responses
CC and cancer and for promoting wound healing. The sequences are used to
CC induce the proliferation of neural cells and regeneration of nerve and
CC brain tissue, and are useful for the treatment of central and peripheral
CC nervous system diseases and neuropathies, such as Alzheimer's disease,
CC Parkinson's disease, Huntington's disease and amyotrophic lateral
CC sclerosis. The sequences are involved in chemotactic or chemokinetic
CC activity, regulation of haematopoiesis, treatment of myeloid or lymphoid
CC cell disorders and platelet disorders such as thrombocytopenia,

CC regeneration of bone, cartilage, tendon, ligament and/or nerve tissue
CC growth, tissue repair, healing of burns, incisions, ulcers, treatment of
CC osteoporosis, osteoarthritis, bone degenerative disorders and periodontal
CC disease. The sequences of the invention are also useful for gut
CC protection or regeneration and treatment of lung or liver fibrosis,
CC reperfusion injury in various tissues, immune deficiencies and disorders
CC including severe combined immunodeficiency (SCID), bacterial or fungal
CC infections, autoimmune disorders e.g. multiple sclerosis and myasthenia
CC gravis, allergic conditions such as asthma, thrombolysis or thrombosis
CC and coagulation disorders. Sequences ABG66666-ABG666758 represent human
CC novel polypeptides of the invention.
XX
SQ Sequence 374 AA;
Query Match 100.0%; Score 1115; DB 23; Length 374;
Best Local Similarity 100.0%; Pred. No. 3.5e-105;
Matches 206; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ATGRLLSGQPVCRGGTQPCYKVIYFHDTSRRLNFEEAKEACRRDGGQLVSI ESEDEQKL 60
|||||
Db 22 ATGRLLSGQPVCRGGTQPCYKVIYFHDTSRRLNFEEAKEACRRDGGQLVSI ESEDEQKL 81
|||||
QY 61 IEKFIENLLPSDGFWIGLRRREKQSNSTACQDLYAWTDGSI SQFRNMYVDEPSCGSEV 120
|||||
Db 82 IEKFIENLLPSDGFWIGLRRREKQSNSTACQDLYAWTDGSI SQFRNMYVDEPSCGSEV 141
|||||
QY 121 CVVMYHOPSAPAGIGGYPYMFQWDDRCNMKNNFICKYSDEKPAVPSREAEGEETELTPV 180
|||||
Db 142 CVVMYHOPSAPAGIGGYPYMFQWDDRCNMKNNFICKYSDEKPAVPSREAEGEETELTPV 201
|||||
QY 181 LPEETQEEDAKKTFKESREAAALNLAY 206
|||||
Db 202 LPEETQEEDAKKTFKESREAAALNLAY 227
|||||
RESULT 4
ABG90203
ID ABB90203 standard; Protein; 374 AA.
XX
AC ABB90203;
XX
DT 24-MAY-2002 (first entry)
XX
DE Human polypeptide SEQ ID NO 2579.
XX
KW Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;
KW antiallergic; hepatotropic; antidiabetic; antiinflammatory; antiulcer;
KW vulnery; anticonvulsant; antibacterial; antifungal; antiparasitic;
KW cardiant; gene therapy; cancer; immune disorder; cardiovascular disorder;
KW neurological disease; infection; human; secreted protein.
XX
OS Homo sapiens.
XX
PN WO200190304-A2.
XX
PD 29-NOV-2001.
XX
PF 18-MAY-2001; 2001WO-US16450.
XX
PR 19-MAY-2000; 2000US-205515P.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Birse CE, Rosen CA;
XX
DR WPI; 2002-122018/16.
DR N-PSDB; ABL90612.
XX
PT Novel 1405 isolated polypeptides, useful for diagnosis, treatment and
PT prevention of neural, immune system, muscular, reproductive,
PT gastrointestinal, pulmonary, cardiovascular, renal and proliferative
PT disorders -
XX

```
PS Claim 11; SEQ ID NO 2579; 208lpp + Sequence Listing; English.
```

```
XX
```

```
CC The invention relates to novel genes (ABL89449-ABL90853) and proteins
```

```
CC (ABB89040-ABB90444) useful for preventing, treating or ameliorating
```

```
CC medical conditions e.g. by protein or gene therapy. The genes are
```

```
CC isolated from a range of human tissues disclosed in the specification.
```

```
CC The nucleic acids, proteins, antibodies and (ant)agonists are useful
```

```
CC in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast
```

```
CC and ovarian cancer and other cancers of the adrenal gland, bone, bone
```

```
CC marrow, breast, gastrointestinal tract, liver, lung, or urogenital;
```

```
CC (b) immune disorders e.g. Addison's disease, allergies, autoimmune
```

```
CC haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's
```

```
CC disease, multiple sclerosis, rheumatoid arthritis and ulcerative
```

```
CC colitis; (c) cardiovascular disorders such as myocardial ischaemias;
```

```
CC (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and
```

```
CC epilepsy; and (f) infectious diseases such as viral, bacterial, fungal
```

```
CC and parasitic infections.
```

```
CC Note: The sequence data for this patent did not form part of the
```

```
CC printed specification, but was obtained in electronic format directly
```

```
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.
```

```
XX
```

```
SQ      Sequence          374 AA;
```

```
Query Match           100.0%; Score 1115; DB 23; Length 374;
```

```
Best Local Similarity 100.0%; Pred. No. 3.5e-105;
```

```
Matches 206; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY    1 ATGRLLSGQPVCVCGGTQRPCYKVIFYHDTSRRLNFEAKEACRRDGGQLVSIESEDEQKL 60  
      |||||||  
Db     22 ATGRLLSGQPVCVCGGTQRPCYKVIFYHDTSRRLNFEAKEACRRDGGQLVSIESEDEQKL 81  
  
QY    61 IEKFIENLLPSDGDFWIGLRRRREEKQSNTACQDLAYATDGSISQFRNWYVDEPSCGSEV 120  
      |||||||  
Db     82 IEKFIENLLPSDGDFWIGLRRRREEKQSNTACQDLAYATDGSISQFRNWYVDEPSCGSEV 141  
  
QY   121 CVVMYHQSPAPAGIGPYMFQWNDDRCNMKNFICKYSDEKPAPVPSREAEGEETELTPV 180  
      |||||||  
Db   142 CVVMYHQSPAPAGIGPYMFQWNDDRCNMKNFICKYSDEKPAPVPSREAEGEETELTPV 201  
  
QY   181 LPEETOEDAKKKTFKRSEAALNLAY 206  
      |||||||  
Db   202 LPEETOEDAKKKTFKRSEAALNLAY 227
```

RESULT 5	
AAM25796	
ID	AAM25796 standard; Protein; 387 AA.
XX	
XX	AAM25796;
XX	
DT	16-OCT-2001 (first entry)
XX	
DE	Human protein sequence SEQ ID NO:1311.
XX	
KW	Human; cancer; ulcer; HIV infection; human immunodeficiency virus;
KW	antiinflammatory; antirheumatic; antiarthritic; immunosuppressive;
KW	antibacterial; endocrine; cardiant; central nervous system; virucide;
KW	anti-HIV; fungicide; antimutagen; cardiovascular; antianaemic; anaemia;
KW	antiaggregant; haemostatic; vulnery; antiulcer; osteopathic; eczema;
KW	dermatological; antiallergic; antiasthmatic; antidiabetic; cytostatic;
KW	neuroprotective; antidepressant; nootropic; antiparkinsonian; infection;
KW	immunostimulant; gene therapy; antisense therapy; septic shock; pancreatitis;
KW	antianaphylactic; rheumatoid arthritis; vaccine; inflammation;
KW	cardiac dysfunction; neuropathology; cardiac anaphylaxis; autoimmunity;
KW	genetic disease; haematopoietic disorder; platelet disorder; asthma;
KW	thrombocytopaenia; osteoporosis; severe combined immunodeficiency;
KW	allergic rhinitis; diabetes; multiple sclerosis; depression;
KW	Alzheimer's disease; Parkinson's disease; neurodegenerative disorder;
XX	neurological disorder.
XX	
OS	Homo sapiens.
XX	
PN	WO200153455-A2.

XX	26-JUL-2001.
PD	
XX	
PF	22-DEC-2000; 2000WO-US35017.
XX	
PR	23-DEC-1999; 99US-0471275.
PR	21-JAN-2000; 2000US-0488725.
PR	25-APR-2000; 2000US-0552317.
XX	
PA	(HYSE-) HYSEQ INC.
XX	
PI	Tang YT, Liu C, Drmanac RT;
XX	
DR	WPI; 2001-457603/49.
DR	N-PSDB; AAH99737.
XX	
PT	Isolated human polynucleotides encoding polypeptides, useful for the
PT	treatment and diagnosis of e.g. cancer, ulcers and HIV infection -
XX	
PS	Claim 20; Page 272; 1217pp; English.
XX	
CC	AAH99166 to AAH99904 encode the human proteins given in AAM25225 to
CC	AAH25963. The proteins can have activities based on the tissues and
CC	cells they are expressed in, such as: antiinflammatory; antirheumatic;
CC	antiarthritic; immunosuppressive; antibacterial; endocrine; cardiac;
CC	central nervous system; virucide; anti-HIV; fungicide; antimutagen;
CC	cardiovascular; antianaemic; antiaggregant; haemostatic; vulnery;
CC	antulcer; osteopathic; dermatological; antiallergic; antiasthmatic;
CC	antidiabetic; cytostatic; neuroprotective; antidepressant; nootropic;
CC	antiparkinsonian; and immunostimulant. The proteins and polynucleotides
CC	encoding them can be used in gene therapy, antisense therapy and vaccine
CC	production. The proteins and polynucleotides are useful for screening for
CC	agonists or antagonists of a protein and for the treatment and diagnosis
CC	of disorders associated with the activity of a protein e.g. inflammation,
CC	rheumatoid arthritis, septic shock, pancreatitis, cardiac dysfunction,
CC	neuropathology, cardiac anaphylaxis, viral, bacterial, HIV and fungal
CC	infections, autoimmunity, genetic diseases, haematopoietic disorders,
CC	anaemia, platelet disorders, thrombocytopaenia, wounds, burns, ulcers,
CC	osteoporosis, severe combined immunodeficiency, eczema, allergic
CC	rinitis, asthma, diabetes, cancer, multiple sclerosis, depression,
CC	Alzheimer's disease, Parkinson's disease, neurodegenerative and
CC	neurological disorders.
XX	
SQ	Sequence 387 AA;

XX DE Human secreted protein sequence encoded by gene 40 SEQ ID NO:163.

XX KW Human; secreted protein; diagnosis; cytostatic; immunosuppressive;

KW antiHIV; antinflammatory; nootropic; neuroprotective; antiallergic;

KW osteopathic; antiarthritic; antibacterial; antidiabetic; antiasthma;

KW antipsoriatic; cardiant; gene therapy; cancer; neurological disorder;

KW immune disease; inflammation; blood disorder; tumour.

XX KW Homo sapiens.

XX OS WO200006698-A1.

XX PN 10-FEB-2000.

PD XX 29-JUL-1999; 99WO-US17130.

PF XX 30-JUL-1998; 98US-0094657.

PR XX 05-AUG-1998; 98US-0095486.

PR XX 06-AUG-1998; 98US-0095454.

PR XX 06-AUG-1998; 98US-0095455.

PR XX 12-AUG-1998; 98US-0096319.

XX PA (HUMA-) HUMAN GENOME SCI INC.

XX PI Komatsoulis GA, Rosen CA, Ruben SM, Duan R, Moore PA, Shi Y;

PI Lafleur D, Wei Y, Ni J, Florence KA, Young PE, Brewer LA;

PI Soppet DR, Endress GA, Ebner R, Olsen HS, Mucenski M;

XX WPI; 2000-195282/17.

DR N-PSDB; AAA26385.

DR XX New isolated human genes and the secreted polypeptides they encode,

PT useful for diagnosis and treatment of e.g. cancers, neurological

PT disorders, immune diseases, inflammation or blood disorders

XX Claim 11; Page 483-484; 634pp; English.

XX CC The polynucleotide sequences given in AAA26346 to AAA26458 encode the

CC human secreted proteins given in AAY91451 to AAY91691. The human secreted

CC proteins can have activities based on the tissues and cells they are

CC expressed in. Examples of the activities are: cytostatic;

CC immunosuppressive; antiHIV; antiinflammatory; nootropic; neuroprotective;

CC antiallergic; osteopathic; antiarthritic; antibacterial; antidiabetic;

CC antiasthma; antipsoriatic; and cardiant. The polynucleotides and their

CC corresponding secreted proteins are useful for preventing, treating or

CC ameliorating medical conditions, e.g. by protein or gene therapy. Also

CC pathological conditions can be diagnosed by determining the amount of the

CC proteins in a sample or by determining the presence of mutations in the

CC polynucleotides. Specific uses are described for each of the

CC polynucleotides, based on which tissues they are most highly expressed

CC in, and include developing products for the diagnosis or treatment of

CC cancer, tumours, neurodegenerative disorders, developmental abnormalities

CC and foetal deficiencies, blood disorders, diseases of the immune system,

CC autoimmune diseases, hepatic and renal disease, inflammation,

CC allergies, Alzheimer's and behavioural disorders, schizophrenia,

CC osteoporosis, arthritis, infections, AIDS, spinal cord injuries,

CC transplant rejection, diabetes, asthma, sepsis, acne, psoriasis,

CC cardiovascular disorders, reproductive disorders, gastrointestinal

CC disorders, respiratory disorders and metabolic disorders. The proteins

CC or polynucleotides can also be used as food additives or preservatives.

CC The proteins are also useful for identifying their binding partners.

CC AAA26337 to AAA26345 and AAY91450 are sequences used in the

CC exemplification of the present invention.

XX SQ Sequence 374 AA;

Query Match 99.1%; Score 1105; DB 21; Length 374;

Best Local Similarity 99.0%; Pred. No. 3.6e-104;

Matches 204; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ATGRLSGQPVCRGQTQPCYKVIYFHDTSRRLNFEAKEACRRDGGQLVSIEDEQKL 60

|||||

Db 22 ATGRLSGQPVCRGQTQPCYKVIYFHDTSRRLNFEAKEACRRDGGQLVSIEDEQKL 81

QY 61 IEKFIENLLPSDGFWIGLRRREKQSNSTACQDLYAWTDGSIQFRNYYVDEPSCGSEV 120

Db 82 IEXFIENLLPSDGFWIGLRRREKQSNSTXCQDLYAWTDGSIQFRNYYVDEPSCGSEV 141

QY 121 CVVMYHQPSAPAGIGGPFYMFQWNDRCNMKNFNICKYSDEKPAVPSREAEGEETELTPV 180

Db 142 CVVMYHQPSAPAGIGGPFYMFQWNDRCNMKNFNICKYSDEKPAVPSREAEGEETELTPV 201

QY 181 LPEETQEEDAKKTFKESREAAALNLAY 206

Db 202 LPEETQEEDAKKTFKESREAAALNLAY 227

RESULT 7

AAAY13367

ID AAY13367 standard; Protein; 382 AA.

XX AC AAY13367;

XX DT 25-JUN-1999 (first entry)

XX DE Amino acid sequence of protein PRO234.

XX KW Secreted protein; transmembrane protein; human; enterocolitis;

KW Zollinger-Ellison syndrome; gastrointestinal ulceration;

KW congenital microvillus atrophy; skin disease; cell growth;

KW abnormal keratinocyte differentiation; psoriasis; epithelial cancer;

KW Parkinson's disease; Alzheimer's disease; ALS; neuropathy;

KW fibromodulin; dermal scarring; Usher Syndrome; Atrophla areata;

KW anti-thrombotic; wound healing; tissue repair.

XX OS Homo sapiens.

XX PN WO9914328-A2.

XX PD 25-MAR-1999.

XX PF 16-SEP-1998; 98WO-US19330.

XX PR 25-NOV-1997; 97US-0066840.

PR 17-SEP-1997; 97US-0059113.

PR 17-SEP-1997; 97US-0059115.

PR 17-SEP-1997; 97US-0059117.

PR 17-SEP-1997; 97US-0059119.

PR 17-SEP-1997; 97US-0059121.

PR 17-SEP-1997; 97US-0059122.

PR 17-SEP-1997; 97US-0059184.

PR 18-SEP-1997; 97US-0059263.

PR 18-SEP-1997; 97US-0059266.

PR 15-OCT-1997; 97US-0062125.

PR 17-OCT-1997; 97US-0062285.

PR 17-OCT-1997; 97US-0062287.

PR 21-OCT-1997; 97US-0063486.

PR 24-OCT-1997; 97US-0062814.

PR 24-OCT-1997; 97US-0062816.

PR 24-OCT-1997; 97US-0063045.

PR 24-OCT-1997; 97US-0063120.

PR 24-OCT-1997; 97US-0063121.

PR 24-OCT-1997; 97US-0063127.

PR 24-OCT-1997; 97US-0063128.

PR 27-OCT-1997; 97US-0063329.

PR 27-OCT-1997; 97US-0063327.

PR 28-OCT-1997; 97US-0063541.

PR 28-OCT-1997; 97US-0063542.

PR 28-OCT-1997; 97US-0063544.

PR 28-OCT-1997; 97US-0063549.

PR 28-OCT-1997; 97US-0063550.

PR 28-OCT-1997; 97US-0063564.

PR 29-OCT-1997; 97US-0063435.

PR 29-OCT-1997; 97US-0063704.

PR 29-OCT-1997; 97US-0063732.

CC activities, and can be used in gene therapy. The PRO polynucleotides,
CC proteins, agonists and antagonists are useful for treating or diagnosing
CC a cardiovascular, endothelial or angiogenic disorder in a mammal,
CC e.g. cardiac hypertrophy, trauma, cancer, age-related macular
CC degeneration, atherosclerosis, hypertension, arterial restenosis,
CC rheumatoid arthritis, angina, myocardial infarctions, thrombophlebitis,
CC lymphangitis, tumour angiogenesis (such as breast carcinoma and liver
CC carcinoma) and wound healing. The PRO polynucleotides have applications
CC in molecular biology, including use as hybridisation probes, and in
CC chromosome and gene mapping. ABL88259 to ABL88267 represent primers and
CC probes used in the exemplification of the present invention.

XX	SQ	Sequence	273 AA;
		Query Match	50.4%; Score 561.5; DB 23; Length 273;
		Best Local Similarity	60.1%; Pred. No. 7e-49;
		Matches	101; Conservative 25; Mismatches 35; Indels 7; Gaps 3;
QY	4	RLLSGQPVCRGGTQPCYKVIYFHDTSRRLNFEEAKEACRRDGGQLVSI ESEDEQKLIK 63	
Db	23	RVVSGQKVCFADFKHPCKYKMAFYHELSSRVSFQEARLACESEGGVLLSLENAEQKLIES 82	
QY	64	FLENLLP-----SDGDFWIGLRRREEKQSNSTACQDLYAWTDGSGISQFRNWIYVDEPSCGS 118	
Db	83	MLQNLTKPGTGISDGDGFWIGLWRNGDGQT-SGACPDLYQWSDGSGNSQYRNWYTDEPSCGS 141	
QY	119	EVCVVMYHQPSAPAGIGGPFYMFQWDDRCNMKNNFICKYSDE-KPAVP 165	
Db	142	EKCVVMYHQPTANPGLGGPYLYQWDDRCNMKNHNYCKYEPEINPTAP 189	

Search completed: May 20, 2003, 17:05:07
Job time : 77 secs

Db	865	ISMCK	-----FIWMDGSKVDFVAWATGEPNFANDDENCVTMY-----TNSGF-----	907
QY	139	MFQWDDRCNMKNFICK	---YSDEKPAVPSPREAGEETEELTTPVLPEETQE-----	187
Db	908	---WINDNCGYPNNEICQRHNSINATAMP-----	TTPTTPGGCKEGWHLYKNK	953
QY	188	-----EDAKKTFKESREALNL	204	
Db	954	CFKIFGFANEKKSWODARQACKGL	978	

RESULT 2
 US-08-840-062-4
 ; Sequence 4, Application US/08840062
 ; Patent No. 6117977
 ; GENERAL INFORMATION:
 ; APPLICANT: LASKY, LAURENCE A.
 ; APPLICANT: WU, KAI
 ; TITLE OF INVENTION: TYPE C LECTINS
 ; NUMBER OF SEQUENCES: 15
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Genentech, Inc.
 ; STREET: 460 Point San Bruno Blvd
 ; CITY: South San Francisco
 ; STATE: California
 ; COUNTRY: USA
 ; ZIP: 94080
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: WinPatIn (Genentech)
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/840,062
 ; FILING DATE:
 ; CLASSIFICATION: 435
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Dreger, Ginger R.
 ; REGISTRATION NUMBER: 33,055
 ; REFERENCE/DOCKET NUMBER: P1019R1
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 415/225-3216
 ; TELEFAX: 415/952-9881
 ; TELEX: 910/371-7168
 ; INFORMATION FOR SEQ ID NO: 4:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 1479 amino acids
 ; TYPE: Amino Acid
 ; TOPOLOGY: Linear
 US-08-840-062-4

		Query Match	15.8%;	Score 176.5;	DB 3;	Length 1479;	
		Best Local Similarity	30.4%;	Pred. NO. 1.4e-09;			
		Matches 51;	Conservative 25;	Mismatches 57;	Indels 35;	Gaps 7;	
<hr/>							
QY	7	SGQPVCRGGTQRPCYKVIYFHDTSRRLLNFEAEKACRRDGGQLVSIESEDEQKLIEKFIE	66				
		: :		:	:	:	:
Db	385	SWQPF-----QGHCYRL-----QA EKRSWQESKKACLRGGDLVSIHSM AELEFITKQIK	434				
		: :		:	:	:	:
<hr/>							
QY	67	NLLPSDGDFFWIGLRRREEKQS NSTACQDLYAWTDGSI SQFRNWYVDEPS---CGSEVCVV	123				
		: : :		:	:	:	:
Db	435	QEVE----ELWIGL-----NDLK LQMNFWS DGSLSFTWHHPPEPNFRDSLED CVT	483				
		: : :		:	:	:	:
<hr/>							
QY	124	MYHQSPAPAGIGPYMFQWNDDRCNMKNFNICKYSDEKPAVPSPREAG	171				
		: : : :		:	:	:	:
Db	484	IW-----GPEG-----RWNDSPCNQSLPSICKKAGQLSQGA A EEDHG	520				
		: : : :		:	:	:	:

RESULT 3
US-08-340-428B-49
; Sequence 49, Application US/08340428B
; Patent No. 5648455
; GENERAL INFORMATION:

```

; APPLICANT: MARGOLIS, Richard U.
; APPLICANT: RAUCH, Uwe
; APPLICANT: MARGOLIS, Renee K.
; TITLE OF INVENTION: CLONING, EXPRESSION AND USES
; TITLE OF INVENTION: NEUROCAN AS A CHONDROITIN SU
; NUMBER OF SEQUENCES: 49
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Browdy and Neimark
; STREET: 419 Seventh Street, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/340,428B
; FILING DATE: 14 No. 5648465ember 1994
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/922,911
; FILING DATE: 03 August 1992
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Browdy, Roger L.
; REGISTRATION NUMBER: 25,618
; REFERENCE/DOCKET NUMBER: Margolis-1A
; TELEPHONE: 202-628-5197
; TELEFAX: 202-737-3528
; INFORMATION FOR SEQ ID NO: 49:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1257 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
;
US-08-340-428B-49

```

```

Query Match      15.6%; Score 174; DB 1; Length 1257;
Best Local Similarity 30.8%; Pred. No. 2e-09;
Matches 44; Conservative 18; Mismatches 49; Indels 32; Gaps 6;

QY 17 QRPCYKVIYFHTSRRLNFEEAKEACRRDGGQLVSIESEDEQKLIKFIENLLPSDGDWF 76
      | ||| : | : ||: ||| | | : | | : | | | : | | : |
Db 1037 QGHCYR--YF---AHRRAWEDAERDCRRRAGHLTSVHSPEEHKFINSF-----GHENSW 1085

QY 77 IGLRRREEKQSNSTACQDLYAWTDGSI SQFRNWWYVDEPS---CGSEVCVVMYHQPSAPAG 133
      ||| | : : : ||| : | | : | | : | | | | : | | | |
Db 1086 IGLNDRITVERD-----FQWTDNTGLQYENWREKQPDNFFAGGEDCVVMVAHENG--- 1134

QY 134 IGGPYMFQWNDRCNMKNNFICK 156
      : ||| || : : ||
Db 1135 -----RWNDVPCNYNLPYVCK 1150

```

RESULT 4
5180808-2
; Patent No. 5180808
; APPLICANT: RUOSLAHTI, ERKKI I.
; TITLE OF INVENTION: VERSICAN CORE PROTEIN, NUCLEIC ACID
; SEQUENCES ENCODING THE SAME, NUCLEIC ACID PROBES, ANTI-VERSICAN
; ANTIBODIES, AND METHODS OF DETECTING THE SAME
; NUMBER OF SEQUENCES: 4
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/441,179
; FILING DATE: 27-NOV-1989
; SEQ ID NO: 2:
; LENGTH: 2409
5180808-2

```

Query Match      15.6%; Score 174; DB 6; Length 2409;
Best Local Similarity 28.5%; Pred. No. 4.9e-09;
Matches 47; Conservative 23; Mismatches 55; Indels 40; Gaps 8;

QY 17 QRPCYKVIYFHTSRLNFEAAKEACRRDGGQLVSISEDEQKLIKFIEILLPSDGF- 75
      | ||| || : | : : || | | | | | : | : | :
Db 2190 QGQCYK--YF---AHRRTWDAARECRQLGAHLTSILSHEEQMFVNRV-----GHDIQ 2237

QY 76 WIGLRRREEKQSNSTACQDLYAWTDGSIQFRNYYVDEP-----SCGSEVCVMYHQPSAP 131
      |||| | : | |||| | : | : | : | : | : | : | : | :
Db 2238 WIGL-----NDKMFHEHDFRWTGDSTLQYENWRPNQPDFSFSAGEDCVVIIWHENG--- 2287

QY 132 AGIGGPMYFQWDDRCNMKNNFICKYS-----DEKPAVPSREAEGE 172
      |||| | | : | | : | | : | | : | : | : | :
Db 2288 -----QWNDVPCNYHLTYTCKKGTVACGQPPVYVENAKTFGK 2323

```


QY	17	QRPCYKVIYFHDTSRRLLNFEEAKEACRRDGGQGLVSI	EESEDEQKLEKFIENLLPSD	GDFW	76
			: :	:	
Db	24	QRKCY---YFGKGTQK--WVHARYACDDMEGQGLVSI	HSPEQDFLTKH-----ASHTGSW	73	
			: :	:	
QY	77	IGLRRREEKQSNSTACQDLYAWTDGSGISQFRNMYVDE	PCSGS--EVCVVMYHQPSAPAGI	134	
			:	: :	
Db	74	IGLRNLDLKE-----FIWVDGSHVDYSNWPCEP	TSRSQGEDCVM-----RGS	118	
			:	: :	
QY	135	GGPYMFQWDDRCNMK--NNEFICKYSDEKPAV---	PSRAEAGE-----ETELTPV	180	
		: : :	: : :		
Db	119	G-----RWNDAFCDRKLGAWVC---DRLATCTPPA	SEGSAESMGPDSPDPDGRLP	TPS	169
			:	:	
QY	181	Lp	182		
Db	170	Ap	171		

```

RESULT 10
US-08-365-103B-10
; Sequence 10, Application US/08365103B
; Patent No. 5766943
; GENERAL INFORMATION:
; APPLICANT: Lynch, Richard G
; APPLICANT: Nunez, Raphael D.
; APPLICANT: Yodoi, Jungi
; TITLE OF INVENTION: DNA Sequences for Soluble Froms of CD23
; TITLE OF INVENTION: and Methods of Use for Same
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Zarley, McKee, Thomte, Voorhees & Sease
; STREET: 801 Grand Ave. Suite 3200
; CITY: Des Moines
; STATE: Iowa
; COUNTRY: United States
; ZIP: 50309
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/365,103B
; FILING DATE: 28-DEC-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Nebel, Heidi S.
; REGISTRATION NUMBER: 37,719
; REFERENCE/DOCKET NUMBER: Uirf N5-24
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (515) 288-3667
; TELEFAX: (515) 288-1338
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 320 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
;
US-08-365-103B-10

```

```

Db 265 G-----RWNDAFCDRKLGAWVC-----DRLATCTPPASEGSAESMGPDSPDPDGRLPDTPS 315

QY 181 LP 182
   |
Db 316 AP 317

RESULT 11
US-08-365-103B-8
; Sequence 8, Application US/08365103B
; Patent No. 5766943
; GENERAL INFORMATION:
; APPLICANT: Lynch, Richard G
; APPLICANT: Nunez, Raphael D.
; APPLICANT: Yodoi, Jungi
; TITLE OF INVENTION: DNA Sequences for Soluble Froms of CD23
; TITLE OF INVENTION: and Methods of Use for Same
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Zarley, McKee, Thomte, Voorhees & Sease
; STREET: 801 Grand Ave. Suite 3200
; CITY: Des Moines
; STATE: Iowa
; COUNTRY: United States
; ZIP: 50309
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/365,103B
; FILING DATE: 28-DEC-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Nebel, Heidi S.
; REGISTRATION NUMBER: 37,719
; REFERENCE/DOCKET NUMBER: Uirf N5-24
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (515) 288-3667
; TELEFAX: (515) 288-1338
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 321 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-365-103B-8

Query Match 13.4%; Score 149; DB 1; Length 321;
Best Local Similarity 30.2%; Pred. No. 1.3e-07;
Matches 55; Conservative 20; Mismatches 57; Indels 50; Gaps 11;

QY 17 QRPCYKVIYFHTSRRLNFEAEKACRRDGGQLVSISEDEQKLIKFIENLLPSDGDWF 76
   || || || : : : || : || || || || : || : |
Db 171 QRKCY--YFGKGTQ--WVHARYACDDMEGQLVSIHSPEEQDFLTKH-----ASHTGSW 220

QY 77 IGLRRREEKQNSTACQDLYAWTDGSIQFRNWWYDEPSCGS--EVCVVMYHQPSAPAGI 134
   || || : | : || : || : || : || : || : |
Db 221 IGLRNLDKGE-----FIWVDGSHVDYSNWAPGEPTSRSQGEDCVNM-----RGS 265

QY 135 GGPYMFQWDDRCNMK--NNFICKYSDEKPAV---PSREAEGE-----ETELTPV 180
   | : || | : | : : | : : | : | : | : |
Db 266 G-----RWNDAFCDRKLGAWVC-----DRLATCTPPASEGSAESMGPDSPDPDGRLPDTPS 316

QY 181 LP 182
   |
Db 317 AP 318

RESULT 12
US-08-751-305-2
; Sequence 2, Application US/08751305

```



```
Query Match      12.9%; Score 143.5; DB 6; Length 372;  
Best Local Similarity   26.5%; Pred. NO. 6e-07;  
Matches    41; Conservative     37; Mismatches     50; Indels     27; Gaps       8;
```

Qy	26 FHDTSRRLLNFEPAKEACRRDGGQLVSISEDEOKLIEKFIIENLLP-SGDGFWIGLRREE 84 : :
Db	41 YHYSEKPMWNENARKFCKQNYTLVAIQNKRE---IE-YLENTLPKSPYYYIWGIRK--- 93 : :
Qy	85 QKSNSTACQDLYAW--TDGSIS-QFRNWYYVDPS--CGSEVCVVMYHQPSAPAGICGPYM 139 : :
Db	94 -----ICKMWTWVGTKTLTKEAENVGAGEPNKKSKEDCVEIYIKRERDSG----- 140 : :
Qy	140 EQWNNDRRCNMKNNFICKYSDEKPVPSPSREAEGEET 174 : :
Db	141 -KWNDADACHKRKAALCYATASCOPGSCNGRGCEVT 174 : :

Search completed: May 20, 2003, 17:07:16
Job time : 17 secs

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 20, 2003, 17:06:10 ; Search time 56 Seconds
(without alignments)
364.814 Million cell updates/sec

Title: US-09-887-855-5
Perfect score: 1115
Sequence: 1 ATGRLLSGQPCVCRGGTQRPC.....EEDAKTFKESREAAALNLAY 206

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 375593 seqs, 99172665 residues

Total number of hits satisfying chosen parameters: 375593

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications_AA:*

- 1: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
- 2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep.*
- 3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep.*
- 4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*
- 5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
- 6: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
- 7: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep.*
- 8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
- 9: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep.*
- 10: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep.*
- 11: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep.*
- 12: /cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB.pep.*
- 13: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep.*
- 14: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1115	100.0	206	10 US-09-887-855-5	Sequence 5, Appli
2	1115	100.0	374	9 US-10-149-819-15	Sequence 15, Appl
3	1115	100.0	374	10 US-09-887-855-2	Sequence 2, Appli
4	1101	98.7	382	9 US-09-905-291A-137	Sequence 137, App
5	1101	98.7	382	9 US-09-902-853-137	Sequence 137, App
6	1101	98.7	382	9 US-09-907-824-137	Sequence 137, App
7	1101	98.7	382	9 US-09-907-841-137	Sequence 137, App
8	1101	98.7	382	9 US-09-904-011-137	Sequence 137, App
9	1101	98.7	382	9 US-10-174-590-20	Sequence 20, Appl
10	1101	98.7	382	9 US-10-176-758-20	Sequence 20, Appl
11	1101	98.7	382	9 US-10-175-737-20	Sequence 20, Appl
12	1101	98.7	382	9 US-09-906-742-137	Sequence 137, App
13	1101	98.7	382	9 US-10-173-706-20	Sequence 20, Appl
14	1101	98.7	382	9 US-10-175-738-20	Sequence 20, Appl
15	1101	98.7	382	9 US-10-175-752-20	Sequence 20, Appl
16	1101	98.7	382	9 US-10-176-482-20	Sequence 20, Appl
17	1101	98.7	382	9 US-10-176-757-20	Sequence 20, Appl
18	1101	98.7	382	9 US-10-176-913-20	Sequence 20, Appl
19	1101	98.7	382	9 US-10-180-552-20	Sequence 20, Appl

20	1101	98.7	382	9 US-10-180-557-20	Sequence 20, Appl
21	1101	98.7	382	9 US-09-906-838-137	Sequence 137, App
22	1101	98.7	382	9 US-09-907-613-137	Sequence 137, App
23	1101	98.7	382	9 US-09-907-942-137	Sequence 137, App
24	1101	98.7	382	9 US-10-173-700-20	Sequence 20, Appl
25	1101	98.7	382	9 US-10-174-572-20	Sequence 20, Appl
26	1101	98.7	382	9 US-10-174-579-20	Sequence 20, Appl
27	1101	98.7	382	9 US-10-174-582-20	Sequence 20, Appl
28	1101	98.7	382	9 US-10-174-588-20	Sequence 20, Appl
29	1101	98.7	382	9 US-10-175-739-20	Sequence 20, Appl
30	1101	98.7	382	9 US-10-175-740-20	Sequence 20, Appl
31	1101	98.7	382	9 US-10-175-743-20	Sequence 20, Appl
32	1101	98.7	382	9 US-10-176-488-20	Sequence 20, Appl
33	1101	98.7	382	9 US-10-176-492-20	Sequence 20, Appl
34	1101	98.7	382	9 US-10-176-747-20	Sequence 20, Appl
35	1101	98.7	382	9 US-10-176-750-20	Sequence 20, Appl
36	1101	98.7	382	9 US-10-176-985-20	Sequence 20, Appl
37	1101	98.7	382	9 US-10-176-987-20	Sequence 20, Appl
38	1101	98.7	382	9 US-10-176-991-20	Sequence 20, Appl
39	1101	98.7	382	9 US-10-176-992-20	Sequence 20, Appl
40	1101	98.7	382	9 US-10-176-993-20	Sequence 20, Appl
41	1101	98.7	382	9 US-10-184-658-20	Sequence 20, Appl
42	1101	98.7	382	9 US-10-173-695-20	Sequence 20, Appl
43	1101	98.7	382	9 US-10-173-697-20	Sequence 20, Appl
44	1101	98.7	382	9 US-10-173-705-20	Sequence 20, Appl
45	1101	98.7	382	9 US-10-174-576-20	Sequence 20, Appl

ALIGNMENTS

RESULT 1
US-09-887-855-5
; Sequence 5, Application US/09887855
; Patent No. US20020058310A1
; GENERAL INFORMATION:
; APPLICANT: Immunex Corporation
; APPLICANT: Anderson, Dirk M
; TITLE OF INVENTION: LECTIN SS3939 DNA AND POLYPEPTIDES
; FILE REFERENCE: 2883-US
; CURRENT APPLICATION NUMBER: US/09/887,855
; CURRENT FILING DATE: 2001-06-22
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 5
; LENGTH: 206
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-887-855-5

Query Match 100.0%; Score 1115; DB 10; Length 206;
Best Local Similarity 100.0%; Pred. No. 6.6e-103;
Matches 206; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	ATGRLLSGQPCVCRGGTQRPCYKVIYFHDTSRRLNFEEAKEACRRDGGQLVSI	60	Sequence 5, Appli
Db	1	ATGRLLSGQPCVCRGGTQRPCYKVIYFHDTSRRLNFEEAKEACRRDGGQLVSI	60	Sequence 2, Appli
QY	61	IEKFIENLLPSDGFWIGLRRREKQSNSTACQDLYAWTDGSIQFRNMYVDEPSCGSEV	120	Sequence 137, App
Db	61	IEKFIENLLPSDGFWIGLRRREKQSNSTACQDLYAWTDGSIQFRNMYVDEPSCGSEV	120	Sequence 137, App
QY	121	CVVMYHOPSAPAGIGGYPYMFQWDDRCNMKNFICKYSDEKPAVPSREAEGETELTPV	180	Sequence 20, Appl
Db	121	CVVMYHOPSAPAGIGGYPYMFQWDDRCNMKNFICKYSDEKPAVPSREAEGETELTPV	180	Sequence 20, Appl
QY	181	LPEETQEEDAKTKFKESREAAALNLAY	206	Sequence 137, App
Db	181	LPEETQEEDAKTKFKESREAAALNLAY	206	Sequence 20, Appl

RESULT 2
US-10-149-819-15


```
; Sequence 15, Application US/10149819
; Publication No. US20030044913A1
; GENERAL INFORMATION:
; APPLICANT: INCYTE GENOMICS, INC.
; APPLICANT: YUE, Henry
; APPLICANT: AZIMZAI, Yalda
; APPLICANT: TANG, Y. Tom
; APPLICANT: PATTERSON, Chandra
; APPLICANT: BAUGHN, Mariah R.
; APPLICANT: LU, Dyung Aina M.
; APPLICANT: SHAH, Purvi
; APPLICANT: LAL, Preeti
; APPLICANT: AU-YOUNG, Janice
; APPLICANT: BURFORD, Neil
; TITLE OF INVENTION: EXTRACELLULAR MATRIX AND CELL ADHESION MOLECULES
; FILE REFERENCE: PF-0760 PCT
; CURRENT APPLICATION NUMBER: US/10/149,819
; PRIOR FILING DATE: 2002-06-10
; PRIOR APPLICATION NUMBER: 60/172,852; 60/172,354
; PRIOR FILING DATE: 1999-12-10; 1999-12-16
; NUMBER OF SEQ ID NOS: 42
; SOFTWARE: PERL Program
; SEQ ID NO 15
; LENGTH: 374
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte ID No. US20030044913A1 3143411CD1
US-10-149-819-15

Query Match      100.0%; Score 1115; DB 9; Length 374;
Best Local Similarity 100.0%; Pred. No. 1.4e-102;
Matches 206; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGRLLSGQPVCRGGTQPCYKVIYFHDTSRRLNFEEAKEACRRDGGQLVSI ESEDEQKL 60
Db 22 ATGRLLSGQPVCRGGTQPCYKVIYFHDTSRRLNFEEAKEACRRDGGQLVSI ESEDEQKL 81

QY 61 IEKFIENLLPSDGFWIGLRRREKQSNSTACQDLYAWTDGSIQFRNMYVDEPSCGSEV 120
Db 82 IEKFIENLLPSDGFWIGLRRREKQSNSTACQDLYAWTDGSIQFRNMYVDEPSCGSEV 141

QY 121 CVVMYHQPSAPAGIGGPPYMFQWDDRCNMKNFNICKYSDEKPAVPSREAEGETELTTPV 180
Db 142 CVVMYHQPSAPAGIGGPPYMFQWDDRCNMKNFNICKYSDEKPAVPSREAEGETELTTPV 201

QY 181 LPEETQEDAKKTFKESREAAALNLAY 206
Db 202 LPEETQEDAKKTFKESREAAALNLAY 227

RESULT 3
US-09-887-855-2
; Sequence 2, Application US/09887855
; Patent No. US20020058310A1
; GENERAL INFORMATION:
; APPLICANT: Immunex Corporation
; APPLICANT: Anderson, Dirk M
; TITLE OF INVENTION: LECTIN SS3939 DNA AND POLYPEPTIDES
; FILE REFERENCE: 2883-US
; CURRENT APPLICATION NUMBER: US/09/887,855
; CURRENT FILING DATE: 2001-06-22
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2
; LENGTH: 374
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-887-855-2

Query Match      100.0%; Score 1115; DB 10; Length 374;
Best Local Similarity 100.0%; Pred. No. 1.4e-102;
```

```
Matches 206; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGRLLSGQPVCRGGTQPCYKVIYFHDTSRRLNFEEAKEACRRDGGQLVSI ESEDEQKL 60
Db 22 ATGRLLSGQPVCRGGTQPCYKVIYFHDTSRRLNFEEAKEACRRDGGQLVSI ESEDEQKL 81

QY 61 IEKFIENLLPSDGFWIGLRRREKQSNSTACQDLYAWTDGSIQFRNMYVDEPSCGSEV 120
Db 82 IEKFIENLLPSDGFWIGLRRREKQSNSTACQDLYAWTDGSIQFRNMYVDEPSCGSEV 141

QY 121 CVVMYHQPSAPAGIGGPPYMFQWDDRCNMKNFNICKYSDEKPAVPSREAEGETELTTPV 180
Db 142 CVVMYHQPSAPAGIGGPPYMFQWDDRCNMKNFNICKYSDEKPAVPSREAEGETELTTPV 201

QY 181 LPEETQEDAKKTFKESREAAALNLAY 206
Db 202 LPEETQEDAKKTFKESREAAALNLAY 227

RESULT 4
US-09-905-291A-137
; Sequence 137, Application US/09905291A
; Patent No. US20020160374A1
; GENERAL INFORMATION:
; APPLICANT: Genentech, Inc.
; APPLICANT: Ashkenazi, Avi
; APPLICANT: Botstein, David
; APPLICANT: Desnoyers, Luc
; APPLICANT: Eaton, Dan L.
; APPLICANT: Ferrara, Napoleone
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Fong, Sherman
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerber, Hanspeter
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, A.
; APPLICANT: Godowski, Paul J.
; APPLICANT: Grimaldi, Christopher J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Hillan, Kenneth, J.
; APPLICANT: Kljavin, Ivar J.
; APPLICANT: Mather, Jennie P.
; APPLICANT: Pan, James
; APPLICANT: Paoni, Nicholas F.
; APPLICANT: Roy, Margaret Ann
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Williams, P. Mickey
; APPLICANT: Wood, William, I.
; TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
; FILE REFERENCE: 10466-14
; CURRENT APPLICATION NUMBER: US/09/905,291A
; CURRENT FILING DATE: 2001-07-12
; PRIOR APPLICATION NUMBER: PCT/US00/04414
; PRIOR FILING DATE: 2000-02-22
; PRIOR APPLICATION NUMBER: US 60/143,048
; PRIOR FILING DATE: 1999-07-07
; PRIOR APPLICATION NUMBER: US 60/145,698
; PRIOR FILING DATE: 1999-07-26
; PRIOR APPLICATION NUMBER: US 60/146,222
; PRIOR FILING DATE: 1999-07-28
; PRIOR APPLICATION NUMBER: PCT/US99/20594
; PRIOR FILING DATE: 1999-09-08
; PRIOR APPLICATION NUMBER: PCT/US99/20944
; PRIOR FILING DATE: 1999-09-13
; PRIOR APPLICATION NUMBER: PCT/US99/21090
; PRIOR FILING DATE: 1999-09-15
; PRIOR APPLICATION NUMBER: PCT/US99/21547
; PRIOR FILING DATE: 1999-09-15
; PRIOR APPLICATION NUMBER: PCT/US99/23089
; PRIOR FILING DATE: 1999-10-05
; PRIOR APPLICATION NUMBER: PCT/US99/28214
```



```
; APPLICANT: Ashkenazi, Avi
; APPLICANT: Botstein, David
; APPLICANT: Desnoyers, Luc
; APPLICANT: Eaton, Dan L.
; APPLICANT: Ferrara, Napoleone
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Fong, Sherman
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerber, Hanspeter
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, A.
; APPLICANT: Godowski, Paul J.
; APPLICANT: Grimaldi, Christopher J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Hillan, Kenneth, J.
; APPLICANT: Kljavin, Ivar J.
; APPLICANT: Mather, Jennie P.
; APPLICANT: Pan, James
; APPLICANT: Paoni, Nicholas F.
; APPLICANT: Roy, Margaret Ann
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Williams, P. Mickey
; APPLICANT: Wood, William, I.
; TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
; TITLE OF INVENTION: Acids Encoding the Same
; FILE REFERENCE: 10466-14
; CURRENT APPLICATION NUMBER: US/09/907,824
; CURRENT FILING DATE: 2001-07-17
; PRIOR APPLICATION NUMBER: 09/665,350
; PRIOR FILING DATE: 2000-09-18
; PRIOR APPLICATION NUMBER: PCT/US00/04414
; PRIOR FILING DATE: 2000-02-22
; PRIOR APPLICATION NUMBER: US 60/143,048
; PRIOR FILING DATE: 1999-07-07
; PRIOR APPLICATION NUMBER: US 60/145,698
; PRIOR FILING DATE: 1999-07-26
; PRIOR APPLICATION NUMBER: US 60/146,222
; PRIOR FILING DATE: 1999-07-28
; PRIOR APPLICATION NUMBER: PCT/US99/20594
; PRIOR FILING DATE: 1999-09-08
; PRIOR APPLICATION NUMBER: PCT/US99/20944
; PRIOR FILING DATE: 1999-09-13
; PRIOR APPLICATION NUMBER: PCT/US99/21090
; PRIOR FILING DATE: 1999-09-15
; PRIOR APPLICATION NUMBER: PCT/US99/21547
; PRIOR FILING DATE: 1999-09-15
; PRIOR APPLICATION NUMBER: PCT/US99/23089
; PRIOR FILING DATE: 1999-10-05
; PRIOR APPLICATION NUMBER: PCT/US99/28214
; PRIOR FILING DATE: 1999-11-29
; PRIOR APPLICATION NUMBER: PCT/US99/28313
; PRIOR FILING DATE: 1999-11-30
; PRIOR APPLICATION NUMBER: PCT/US99/28564
; PRIOR FILING DATE: 1999-12-02
; PRIOR APPLICATION NUMBER: PCT/US99/28565
; PRIOR FILING DATE: 1999-12-02
; PRIOR APPLICATION NUMBER: PCT/US99/30095
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: PCT/US99/30911
; PRIOR FILING DATE: 1999-12-20
; PRIOR APPLICATION NUMBER: PCT/US99/30999
; PRIOR FILING DATE: 1999-12-20
; PRIOR APPLICATION NUMBER: PCT/US00/00219
; PRIOR FILING DATE: 2000-01-05
; NUMBER OF SEQ ID NOS: 423
; SEQ ID NO 137
; LENGTH: 382
; TYPE: PRT
; ORGANISM: Homo Sapien
US-09-907-824-137
```

Query Match 98.7%; Score 1101; DB 9; Length 382;

```
Best Local Similarity 96.3%; Pred. No. 3.6e-101;
Matches 206; Conservative 0; Mismatches 0; Indels 8; Gaps 1;

QY 1 ATGRLLS-----GQPVCRGGTQRPCYKVIYFHDTSRRLNFEEAKEACRRDGGQLVSI 52
      |||||
Db 22 ATGRLLSASDLRLRGQPVCRGGTQRPCYKVIYFHDTSRRLNFEEAKEACRRDGGQLVSI 81

QY 53 ESEDEQKLIKFIENLLPSDGDWIGLRRRREKOSNSTACQDLYAWTDGSGISOFRNWYVD 112
      |||||
Db 82 ESEDEQKLIKFIENLLPSDGDWIGLRRRREKOSNSTACQDLYAWTDGSGISOFRNWYVD 141

QY 113 EPSCGSEVCVMYHOPSAPAGIGGYPYMFQWDDRCNMKNNFICKYSDEKPAVPSREAEGE 172
      |||||
Db 142 EPSCGSEVCVMYHOPSAPAGIGGYPYMFQWDDRCNMKNNFICKYSDEKPAVPSREAEGE 201

QY 173 ETELTPVLPEETQEDAKKTFKESREAAALNLAY 206
      |||||
Db 202 ETELTPVLPEETQEDAKKTFKESREAAALNLAY 235

RESULT 7
US-09-907-841-137
; Sequence 137, Application US/09907841
; Publication No. US20020198366A1
; GENERAL INFORMATION:
; APPLICANT: Genentech, Inc.
; APPLICANT: Ashkenazi, Avi
; APPLICANT: Botstein, David
; APPLICANT: Desnoyers, Luc
; APPLICANT: Eaton, Dan L.
; APPLICANT: Ferrara, Napoleone
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Fong, Sherman
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerber, Hanspeter
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, A.
; APPLICANT: Godowski, Paul J.
; APPLICANT: Grimaldi, Christopher J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Hillan, Kenneth, J.
; APPLICANT: Kljavin, Ivar J.
; APPLICANT: Mather, Jennie P.
; APPLICANT: Pan, James
; APPLICANT: Paoni, Nicholas F.
; APPLICANT: Roy, Margaret Ann
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Williams, P. Mickey
; APPLICANT: Wood, William, I.
; TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
; TITLE OF INVENTION: Acids Encoding the Same
; FILE REFERENCE: 10466-14
; CURRENT APPLICATION NUMBER: US/09/907,841
; CURRENT FILING DATE: 2001-11-20
; PRIOR APPLICATION NUMBER: PCT/US00/04414
; PRIOR FILING DATE: 2000-02-22
; PRIOR APPLICATION NUMBER: US 60/143,048
; PRIOR FILING DATE: 1999-07-07
; PRIOR APPLICATION NUMBER: US 60/145,698
; PRIOR FILING DATE: 1999-07-26
; PRIOR APPLICATION NUMBER: US 60/146,222
; PRIOR FILING DATE: 1999-07-28
; PRIOR APPLICATION NUMBER: PCT/US99/20594
; PRIOR FILING DATE: 1999-09-08
; PRIOR APPLICATION NUMBER: PCT/US99/20944
; PRIOR FILING DATE: 1999-09-13
; PRIOR APPLICATION NUMBER: PCT/US99/21090
; PRIOR FILING DATE: 1999-09-15
; PRIOR APPLICATION NUMBER: PCT/US99/21547
; PRIOR FILING DATE: 1999-09-15
; PRIOR APPLICATION NUMBER: PCT/US99/23089
; PRIOR FILING DATE: 1999-10-05
```

```
; PRIOR APPLICATION NUMBER: PCT/US99/28214
; PRIOR FILING DATE: 1999-11-29
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 423
; SEQ ID NO 137
; LENGTH: 382
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-907-841-137

Query Match      98.7%; Score 1101; DB 9; Length 382;
Best Local Similarity 96.3%; Pred. No. 3.6e-101;
Matches 206; Conservative 0; Mismatches 0; Indels 8; Gaps 1;

QY 1 ATGRLLS-----GQVCRGGTQRPCYKVIYFHDTSRRLNFEEAKEACRRDGGQLVSI 52
Db 22 ATGRLLSASDLRLGGQPVCRGGTQRPCYKVIYFHDTSRRLNFEEAKEACRRDGGQLVSI 81

QY 53 ESEDEQKLIKFIENLLPSDGFWIGLRRREEKQSNSTACQDLYAWTDGSGISQFRNWIYVD 112
Db 82 ESEDEQKLIKFIENLLPSDGFWIGLRRREEKQSNSTACQDLYAWTDGSGISQFRNWIYVD 141

QY 113 EPSCGSEVCVMYHQPSAPAGIGGPPYMFQWNDRCNMKNFICKYSDEKPAVPSREAEGE 172
Db 142 EPSCGSEVCVMYHQPSAPAGIGGPPYMFQWNDRCNMKNFICKYSDEKPAVPSREAEGE 201

QY 173 ETELTPVLPEETQEEADAKKTFKESREAAALNLAY 206
Db 202 ETELTPVLPEETQEEADAKKTFKESREAAALNLAY 235

RESULT 8
US-09-904-011-137
; Sequence 137, Application US/09904011
; Publication No. US2003003530A1
; GENERAL INFORMATION:
; APPLICANT: Genentech, Inc.
; APPLICANT: Ashkenazi, Avi
; APPLICANT: Botstein, David
; APPLICANT: Desnoyers, Luc
; APPLICANT: Eaton, Dan L.
; APPLICANT: Ferrara, Napoleone
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Fong, Sherman
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerber, Hanspeter
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, A.
; APPLICANT: Godowski, Paul J.
; APPLICANT: Grimaldi, Christopher J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Hillan, Kenneth, J.
; APPLICANT: Kljavin, Ivar J.
; APPLICANT: Mather, Jennie P.
; APPLICANT: Pan, James
; APPLICANT: Paoni, Nicholas F.
; APPLICANT: Roy, Margaret Ann
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Williams, P. Mickey
; APPLICANT: Wood, William, I.
; TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
; TITLE OF INVENTION: Acids Encoding the Same
; FILE REFERENCE: 10466-14
; CURRENT APPLICATION NUMBER: US/09/904,011
; CURRENT FILING DATE: 2001-07-11
; PRIOR APPLICATION NUMBER: 09/665,350
; PRIOR FILING DATE: 2000-09-18
; PRIOR APPLICATION NUMBER: PCT/US00/04414
; PRIOR FILING DATE: 2000-02-22
; PRIOR APPLICATION NUMBER: US 60/143,048
; PRIOR FILING DATE: 1999-07-07
; PRIOR APPLICATION NUMBER: US 60/145,698
```

```
; PRIOR FILING DATE: 1999-07-26
; PRIOR APPLICATION NUMBER: US 60/146,222
; PRIOR FILING DATE: 1999-07-28
; PRIOR APPLICATION NUMBER: PCT/US99/20594
; PRIOR FILING DATE: 1999-09-08
; PRIOR APPLICATION NUMBER: PCT/US99/20944
; PRIOR FILING DATE: 1999-09-13
; PRIOR APPLICATION NUMBER: PCT/US99/21090
; PRIOR FILING DATE: 1999-09-15
; PRIOR APPLICATION NUMBER: PCT/US99/21547
; PRIOR FILING DATE: 1999-09-15
; PRIOR APPLICATION NUMBER: PCT/US99/23089
; PRIOR FILING DATE: 1999-10-05
; PRIOR APPLICATION NUMBER: PCT/US99/28214
; PRIOR FILING DATE: 1999-11-29
; PRIOR APPLICATION NUMBER: PCT/US99/28313
; PRIOR FILING DATE: 1999-11-30
; PRIOR APPLICATION NUMBER: PCT/US99/28564
; PRIOR FILING DATE: 1999-12-02
; PRIOR APPLICATION NUMBER: PCT/US99/28565
; PRIOR FILING DATE: 1999-12-02
; PRIOR APPLICATION NUMBER: PCT/US99/30095
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: PCT/US99/30911
; PRIOR FILING DATE: 1999-12-20
; PRIOR APPLICATION NUMBER: PCT/US99/30999
; PRIOR FILING DATE: 1999-12-20
; PRIOR APPLICATION NUMBER: PCT/US00/00219
; PRIOR FILING DATE: 2000-01-05
; NUMBER OF SEQ ID NOS: 423
; SEQ ID NO 137
; LENGTH: 382
; TYPE: PRT
; ORGANISM: Homo Saplen
US-09-904-011-137

Query Match      98.7%; Score 1101; DB 9; Length 382;
Best Local Similarity 96.3%; Pred. No. 3.6e-101;
Matches 206; Conservative 0; Mismatches 0; Indels 8; Gaps 1;

QY 1 ATGRLLS-----GQVCRGGTQRPCYKVIYFHDTSRRLNFEEAKEACRRDGGQLVSI 52
Db 22 ATGRLLSASDLRLGGQPVCRGGTQRPCYKVIYFHDTSRRLNFEEAKEACRRDGGQLVSI 81

QY 53 ESEDEQKLIKFIENLLPSDGFWIGLRRREEKQSNSTACQDLYAWTDGSGISQFRNWIYVD 112
Db 82 ESEDEQKLIKFIENLLPSDGFWIGLRRREEKQSNSTACQDLYAWTDGSGISQFRNWIYVD 141

QY 113 EPSCGSEVCVMYHQPSAPAGIGGPPYMFQWNDRCNMKNFICKYSDEKPAVPSREAEGE 172
Db 142 EPSCGSEVCVMYHQPSAPAGIGGPPYMFQWNDRCNMKNFICKYSDEKPAVPSREAEGE 201

QY 173 ETELTPVLPEETQEEADAKKTFKESREAAALNLAY 206
Db 202 ETELTPVLPEETQEEADAKKTFKESREAAALNLAY 235

RESULT 9
US-10-174-590-20
; Sequence 20, Application US/10174590
; Publication No. US20030008352A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jian
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Smith, Victoria
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
```


APPLICANT: Fong, Sherman
APPLICANT: Gao, Wei-Qiang
APPLICANT: Gerber, Hanspeter
APPLICANT: Gerritsen, Mary E.
APPLICANT: Goddard, A.
APPLICANT: Godowski, Paul J.
APPLICANT: Grimaldi, Christopher J.
APPLICANT: Gurney, Austin L.
APPLICANT: Hillan, Kenneth, J.
APPLICANT: Kljavin, Ivar J.
APPLICANT: Mather, Jennie P.
APPLICANT: Pan, James
APPLICANT: Paoni, Nicholas F.
APPLICANT: Roy, Margaret Ann
APPLICANT: Stewart, Timothy A.
APPLICANT: Tumas, Daniel
APPLICANT: Williams, P. Mickey
APPLICANT: Wood, William, I.
TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
TITLE OF INVENTION: Acids Encoding the Same
FILE REFERENCE: 10466-14
CURRENT APPLICATION NUMBER: US/09/906,742
CURRENT FILING DATE: 2001-07-16
PRIOR APPLICATION NUMBER: 09/665,350
PRIOR FILING DATE: 2000-09-18
PRIOR APPLICATION NUMBER: PCT/US00/04414
PRIOR FILING DATE: 2000-02-22
PRIOR APPLICATION NUMBER: US 60/143,048
PRIOR FILING DATE: 1999-07-07
PRIOR APPLICATION NUMBER: US 60/145,698
PRIOR FILING DATE: 1999-07-26
PRIOR APPLICATION NUMBER: US 60/146,222
PRIOR FILING DATE: 1999-07-28
PRIOR APPLICATION NUMBER: PCT/US99/20594
PRIOR FILING DATE: 1999-09-08
PRIOR APPLICATION NUMBER: PCT/US99/20944
PRIOR FILING DATE: 1999-09-13
PRIOR APPLICATION NUMBER: PCT/US99/21090
PRIOR FILING DATE: 1999-09-15
PRIOR APPLICATION NUMBER: PCT/US99/21547
PRIOR FILING DATE: 1999-09-15
PRIOR APPLICATION NUMBER: PCT/US99/23089
PRIOR FILING DATE: 1999-10-05
PRIOR APPLICATION NUMBER: PCT/US99/28214
PRIOR FILING DATE: 1999-11-29
PRIOR APPLICATION NUMBER: PCT/US99/28313
PRIOR FILING DATE: 1999-11-30
PRIOR APPLICATION NUMBER: PCT/US99/28564
PRIOR FILING DATE: 1999-12-02
PRIOR APPLICATION NUMBER: PCT/US99/28565
PRIOR FILING DATE: 1999-12-02
PRIOR APPLICATION NUMBER: PCT/US99/30095
PRIOR FILING DATE: 1999-12-16
PRIOR APPLICATION NUMBER: PCT/US99/30911
PRIOR FILING DATE: 1999-12-20
PRIOR APPLICATION NUMBER: PCT/US99/30999
PRIOR FILING DATE: 1999-12-20
PRIOR APPLICATION NUMBER: PCT/US00/00219
PRIOR FILING DATE: 2000-01-05
NUMBER OF SEQ ID NOS: 423
SEQ ID NO 137
LENGTH: 382
TYPE: PRT
ORGANISM: Homo Sapien
US-09-906-742-137

Query Match 98.7%; Score 1101; DB 9; Length 382;
Best Local Similarity 96.3%; Pred. No. 3.6e-101;
Matches 206; Conservative 0; Mismatches 0; Indels 8; Gaps 1;
Qy 1 ATGRLLS-----GQVCRGGTQPCYKVIYFHDTSRRLNFEEAKEACRRDGGQLVSI 52
Db 22 ATGRLLSASDLRLGGQPCVCRGGTQPCYKVIYFHDTSRRLNFEEAKEACRRDGGQLVSI 81

QY 53 ESEDEQKLIKFIENLLPSDGDGFWIGLRRREKQSNSTACQDLYAWTDGSIQSFRNYYVD 112
Db 82 ESEDEQKLIKFIENLLPSDGDGFWIGLRRREKQSNSTACQDLYAWTDGSIQSFRNYYVD 141
QY 113 EPSCGSEVCVVMYHQPSAPAGIGGPPYMFQWDDRCNMKNNFICKYSDEKPAVPSREAEGE 172
Db 142 EPSCGSEVCVVMYHQPSAPAGIGGPPYMFQWDDRCNMKNNFICKYSDEKPAVPSREAEGE 201
QY 173 ETELTPVLPEETQEEDAKKTFKESREAAALNLAY 206
Db 202 ETELTPVLPEETQEEDAKKTFKESREAAALNLAY 235
RESULT 13
US-10-173-706-20
; Sequence 20, Application US/10173706
; Publication No. US20030022293A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jian
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Smith, Victoria
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; TITLE OF INVENTION: ACIDS ENCODING THE SAME
; FILE REFERENCE: P3430R1C7
; CURRENT APPLICATION NUMBER: US/10/173,706
; CURRENT FILING DATE: 2002-06-17
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 20
; LENGTH: 382
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-173-706-20

Query Match 98.7%; Score 1101; DB 9; Length 382;
Best Local Similarity 96.3%; Pred. No. 3.6e-101;
Matches 206; Conservative 0; Mismatches 0; Indels 8; Gaps 1;
QY 1 ATGRLLS-----GQVCRGGTQPCYKVIYFHDTSRRLNFEEAKEACRRDGGQLVSI 52
Db 22 ATGRLLSASDLRLGGQPCVCRGGTQPCYKVIYFHDTSRRLNFEEAKEACRRDGGQLVSI 81
QY 53 ESEDEQKLIKFIENLLPSDGDGFWIGLRRREKQSNSTACQDLYAWTDGSIQSFRNYYVD 112
Db 82 ESEDEQKLIKFIENLLPSDGDGFWIGLRRREKQSNSTACQDLYAWTDGSIQSFRNYYVD 141
QY 113 EPSCGSEVCVVMYHQPSAPAGIGGPPYMFQWDDRCNMKNNFICKYSDEKPAVPSREAEGE 172
Db 142 EPSCGSEVCVVMYHQPSAPAGIGGPPYMFQWDDRCNMKNNFICKYSDEKPAVPSREAEGE 201
QY 173 ETELTPVLPEETQEEDAKKTFKESREAAALNLAY 206
Db 202 ETELTPVLPEETQEEDAKKTFKESREAAALNLAY 235

RESULT 14
US-10-175-738-20
; Sequence 20, Application US/10175738
; Publication No. US20030022294A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jian
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 20, 2003, 17:03:25 ; Search time 44 Seconds
(without alignments)
450.084 Million cell updates/sec

Title: US-09-887-855-5
Perfect score: 1115
Sequence: 1 ATGRLLSGQPVCRGGTQRPC.....EEDAKTKFKESREALNLAY 206

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues
Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_73:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	182	16.3	1456	1 A36563	mannose receptor p
2	178.5	16.0	1455	1 A48925	mannose receptor p
3	177	15.9	1268	2 S52781	neurocan - mouse
4	174.5	15.7	1643	2 T14274	versican precursor
5	174.5	15.7	3381	2 T42389	versican precursor
6	174	15.6	1257	2 S28764	neurocan precursor
7	174	15.6	2397	1 A55535	versican precursor
8	174	15.6	2409	1 A60979	versican precursor
9	171	15.3	3562	2 A47171	chondroitin sulfat
10	170.5	15.3	1479	2 T42710	mannose receptor,
11	158.5	14.2	1340	2 A39808	proteoglycan core
12	158.5	14.2	2327	2 T42630	aggrecan - bovine
13	158.5	14.2	2415	1 A39086	aggrecan precursor
14	154.5	13.9	612	2 B42755	E-selectin precurs
15	153.5	13.8	2124	2 A28452	proteoglycan core
16	152	13.6	912	2 A54423	brevican precursor
17	149.5	13.4	459	2 T24425	hypothetical prote
18	149	13.4	321	1 LNHUER	IgE Fc receptor II
19	148.5	13.3	330	2 T46256	brevican - human (
20	148.5	13.3	2132	1 A55182	aggrecan precursor
21	148	13.3	253	2 E89130	protein F52E1.2 [1
22	147	13.2	883	2 S57653	brevican precursor
23	146.5	13.1	162	1 LNRCL	lectin BRA3-1 prec
24	146	13.1	2109	1 I50421	aggrecan precursor
25	145.5	13.0	742	2 JC7595	scavenger receptor
26	145	13.0	883	2 S49126	brevican precursor
27	144.5	13.0	173	2 S10548	lectin - barnacle
28	144.5	13.0	372	2 S23936	L-selectin precurs
29	144.5	13.0	404	2 A46274	HIV gp120-binding

30 143.5 12.9 129 2 JC4329 coagulation factor
31 143.5 12.9 372 1 A32375 L-selectin precurs
32 143.5 12.9 463 2 T26655 hypothetical prote
33 142.5 12.8 131 2 JC5058 bitiscetin alpha c
34 142 12.7 1487 2 S48719 phospholipase-A(2)
35 141.5 12.7 331 1 LNMSEr IGE Fc receptor, 1
36 140.5 12.6 162 1 LNRSCr lectin BRA3-2 prec
37 140 12.6 370 2 S22124 L-selectin precurs
38 139 12.5 248 1 LNHUPS pulmonary surfacta
39 139 12.5 248 1 LNHUP6 pulmonary surfacta
40 139 12.5 248 1 LNHUP1 pulmonary surfacta
41 139 12.5 283 1 LNFHLS lectin precursor -
42 138.5 12.4 152 2 JC4690 coagulation factor
43 138.5 12.4 202 2 JC4031 tetranectin precur
44 138 12.4 280 2 T29200 hypothetical prote
45 137.5 12.3 309 1 S34198 IGE Fc receptor II

ALIGNMENTS

RESULT 1
A36563
mannose receptor precursor - human
C;Species: Homo sapiens (man)
C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
C;Accession: A36563; A60926; A44255; B44255; C44255; D44255; E44255; F44255; G44255
R;Taylor, M.E.; Conary, J.T.; Lennartz, M.R.; Stahl, P.D.; Drickamer, K.
J. Biol. Chem. 265, 12156-12162, 1990
A;Title: Primary structure of the mannose receptor contains multiple motifs resembl
A;Reference number: A36563; MUID:90324192; PMID:2373685
A;Accession: A36563
A;Molecule type: mRNA
A;Residues: 1-1456 <TAY>
A;Cross-references: GB:J05550; NID:g188675; PIDN:AAA59868.1; PID:g188676
A;Note: parts of this sequence, including the amino end of the mature protein, were
R;Ezekowitz, R.A.B.; Sastry, K.; Bailly, P.; Warner, A.
J. Exp. Med. 172, 1785-1794, 1990
A;Title: Molecular characterization of the human macrophage mannose receptor: demon
A;Reference number: A60926; MUID:91079783; PMID:2258707
A;Accession: A60926
A;Status: nucleic acid sequence not shown
A;Molecule type: mRNA
A;Residues: 1-1333,'T',1335-1456 <EZE>
A;Cross-references: GB:X55635
A;Note: translation of the nucleotide sequence is incomplete
A;Note: in the authors' translation additional residues Pro-Glu-Ile are shown after
R;Kim, S.J.; Ruiz, N.; Bezouska, K.; Drickamer, K.
Genomics 14, 721-727, 1992
A;Title: Organization of the gene encoding the human macrophage mannose receptor (M
A;Reference number: A44255; MUID:93052405; PMID:1294118
A;Accession: A44255
A;Status: preliminary; nucleic acid sequence not shown; not compared with conceptua
A;Molecule type: DNA
A;Residues: 155-233,'KSAL',238-283;346-428;492-569;631-714,716-719;783-820,'N',822-
A;Note: sequence extracted from NCBI backbone (NCBIP:118415, NCBIP:118421, NCBIP:11
C;Genetics:
A;Gene: GDB:MRC1
A;Cross-references: GDB:133759; OMIM:153618
A;Map position: 10p13-10p13
C;Superfamily: phospholipase A2 receptor; C-type lectin homology; fibronectin type
C;Keywords: duplication; lectin; tandem repeat; transmembrane protein
F;1-18/Domain: signal sequence #status predicted <SIG>
F;168-209/Domain: fibronectin type II repeat homology <2F1>
F;223-340/Domain: C-type lectin homology <LCH1>
F;362-486/Domain: C-type lectin homology <LCH2>
F;945-1079/Domain: C-type lectin homology <LCH3>

Query Match 16.3%; Score 182; DB 1; Length 1456;
Best Local Similarity 25.5%; Pred. No. 9.3e-08;
Matches 52; Conservative 37; Mismatches 61; Indels 54; Gaps 8;

QY 21 YKVIYFHDTSRRLLNFEEAKEACRRDGGQLVSI ESEDEQKLIKFIENLLPSDGFWIGLR 80

Db 807 YKDYQYFSKEKETMDNARAFCKRNFGLVSIQSESKFLWKYV-NRNDQAAYFIGLL 865
QY 81 RREEKQSNSTACQDLYAWTDGSIQFRNWYVDEPSCGS--EVCVVMYHQPSAPAGIGGPY 138
Db 866 ISLDKK-----FAWMDGSKVDYVSWATGEPNFANEDENCVTMY----SNSGF----- 908
QY 139 MFQWDDRCNMKNNFICKYSDEK----PAVPVSREAEGEETELTTPVLPEETQE----- 187
Db 909 ---WNDINGYPNAFICQRHNSINATTVM-----TMSVPSCGCKEGWNFYSN 954
QY 188 -----EDAKKTFKESREAAAL 202
Db 955 KCFKIFGFMEERKNWQEARKACI 978

RESULT 2
A48925
mannose receptor precursor, macrophage - mouse
C:Species: Mus musculus (house mouse)
C>Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
C:Accession: A48925; S21320; PC2245
R:Harris, N.; Super, M.; Rits, M.; Chang, G.; Ezekowitz, R.A.
Blood 80, 2363-2373, 1992
A:Title: Characterization of the murine macrophage mannose receptor: demonstration that
A:Reference number: A48925; MUID:93043353; PMID:1421407
A:Accession: A48925
A>Status: not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 1-1455 <HAR>
A:Experimental source: peritoneal macrophage
A>Note: sequence extracted from NCBI backbone (NCBIP:118733)
R:Harris, N.; Super, M.; Rits, M.; Chang, G.; Ezekowitz, R.B.
Submitted to the EMBL Data Library, April 1992
A:Description: Characterization of the murine macrophage mannose receptor: Demonstration
on.
A:Reference number: S21320
A:Accession: S21320
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-302,'W',303-1117,'E',1119-1455 <HA2>
A:Cross-references: EMBL:Z11974; NID:952997; PIDN:CAA78028.1; PID:952998
R:Harris, N.; Peters, L.L.; Eicher, E.M.; Rits, M.; Raspberry, D.; Eichbaum, Q.G.; Super
Biochem. Biophys. Res. Commun. 198, 682-692, 1994
A:Title: The exon-intron structure and chromosomal localization of the mouse macrophage
A:Reference number: PC2245; MUID:94128116; PMID:8297379
A:Accession: PC2245
A:Molecule type: mRNA
A:Residues: 35-105 <HA3>
C:Genetics:
A:Gene: Mrcl
A:Map position: 2
C:Superfamily: phospholipase A2 receptor; C-type lectin homology; fibronectin type II re
C:Keywords: membrane protein; receptor
F:168-209/Domain: fibronectin type II repeat homology <2F9>
F:361-485/Domain: C-type lectin homology <LCH1>
F:943-1077/Domain: C-type lectin homology <LCH2>

Query Match 16.0%; Score 178.5; DB 1; Length 1455;
Best Local Similarity 25.9%; Pred. No. 1.9e-07;
Matches 53; Conservative 33; Mismatches 66; Indels 53; Gaps 8;

QY 21 YKVIYFHDTSRLNFEAEKACRRDGGQLVSISEDEQKLIKFIENLLPSDGDWIGLR 80
Db 806 YKDYQYFSKEKETMDNARRFCKKNFGDLATIKSESKFLWKYI-NKNGGQSPYFIGML 864
QY 81 RREEKQSNSTACQDLYAWTDGSIQFRNWYVDEPSCGS--EVCVVMYHQPSAPAGIGGPY 138
Db 865 ISMDKK-----FIWMDGSKVDYFVAWATGEPNFANDDENCVTMY----TNSGF----- 907
QY 139 MFQWDDRCNMKNNFICK---YSDERKPAVPVSREAEGEETELTTPVLPEETQE----- 187
Db 908 ---WNDINGYPNNFICQRHNSINATAMP-----TTPTPGGCKEGWHLYKNK 953

QY 188 -----EDAKKTFKESREAAALNL 204
Db 954 CFKIFGFANEKKSWQDARQACKGL 978
RESULT 3
S52781
neurocan - mouse
C:Species: Mus musculus (house mouse)
C>Date: 19-May-1995 #sequence_revision 21-Jul-1995 #text_change 04-Feb-2000
C:Accession: S52781
R:Rauch, U.; Forsberg, N.; Kulbe, G.; Arnold-Ammer, I.; Faessler, R.
submitted to the EMBL Data Library, February 1995
A:Description: Amino acid sequence of mouse neurocan and brevican and their differ
A:Reference number: S52781
A:Accession: S52781
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-1268 <RAU>
A:Cross-references: EMBL:X84727; NID:9758629; PIDN:CAA59216.1; PID:9758630
C:Superfamily: aggrecan; C-type lectin homology; complement factor H repeat homology
F:176-253/Domain: link protein repeat homology <LNK1>
F:274-355/Domain: link protein repeat homology <LNK2>
F:964-995/Domain: EGF homology <EGF>
F:1040-1160/Domain: C-type lectin homology <LCH>
F:1167-1223/Domain: complement factor H repeat homology <FHD>

Query Match 15.9%; Score 177; DB 2; Length 1268;
Best Local Similarity 31.5%; Pred. No. 2.2e-07;
Matches 45; Conservative 17; Mismatches 49; Indels 32; Gaps 6;

QY 17 QRPCYKVIYFHDTSRLNFEAEKACRRDGGQLVSISEDEQKLIKFIENLLPSDGDW 76
Db 1048 QGHCYR--YF---AHRRAWEDAERDCRRRAGHLTSHVSPSEHKFINSF-----GHENSW 1096
QY 77 IGLRRREEKQSNSTACQDLYAWTDGSIQFRNWYVDEPS---CGSEVCVVMYHQPSAPAG 133
Db 1097 IGLNDRTVERD-----FQWTDNTGLQYENWRKQPDNFFAGGEDCVMVMAHESG--- 1145
QY 134 IGGPYMFQWDDRCNMKNNFICK 156
Db 1146 -----RWNDVPCVNYNLPYVCK 1161

RESULT 4
T14274
versican precursor, splice form V2 - bovine
C:Species: Bos primigenius taurus (cattle)
C>Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 05-May-2000
C:Accession: T14274
R:Schmalfeldt, M.; Dours-Zimmermann, M.T.; Winterhalter, K.H.; Zimmermann, D.R.
J. Biol. Chem. 273, 15758-15764, 1998
A:Title: Versican V2 is a major extracellular matrix component of the mature bovine
A:Reference number: Z17954; MUID:98288320; PMID:9624174
A:Accession: T14274
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-1643 <SCH>
A:Cross-references: EMBL:AF060458; NID:93253303; PID:g3253304; PIDN:AAC24360.1
A:Experimental source: brain
C:Keywords: glycoprotein
F:1-20/Domain: signal sequence #status predicted <SIG>
F:21-1643/Product: versican, splice form V2 #status predicted <MAT>
F:57,331,352,817,965,1017,1333,1616,1626/Binding site: carbohydrate (Asn) (covalent)

Query Match 15.7%; Score 174.5; DB 2; Length 1643;
Best Local Similarity 25.5%; Pred. No. 4.9e-07;
Matches 59; Conservative 26; Mismatches 71; Indels 75; Gaps 11;

QY 17 QRPCYKVIYFHDTSRLNFEAEKACRRDGGQLVSISEDEQKLIKFIENLLPSDGD 75
Db 1424 QGQCYK--YF---AHRRTWDAARECRLOQAHLTLSILSHEEQMFVNRV-----GHDYQ 1471

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 20, 2003, 17:00:15 ; Search time 13 Seconds
(without alignments)
657.240 Million cell updates/sec

Title: US-09-887-855-5
Perfect score: 1115
Sequence: 1 ATGRLLSGQPVCRGTQRPC.....EEDAKKTFKESREALNLAY 206

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	574	51.5	273	1	CHOD_MOUSE
2	561.5	50.4	273	1	CHOD_HUMAN
3	182	16.3	1456	1	MANR_HUMAN
4	177	15.9	1268	1	PGCN_MOUSE
5	174.5	15.7	3381	1	PGCV_BOVIN
6	174	15.6	1257	1	PGCN_RAT
7	174	15.6	2738	1	PGCV_RAT
8	174	15.6	3358	1	PGCV_MOUSE
9	174	15.6	3396	1	PGCV_HUMAN
10	171	15.3	3562	1	PGCV_CHICK
11	165	14.8	643	1	CD93_RAT
12	158.5	14.2	2364	1	PGCA_BOVIN
13	158.5	14.2	2415	1	PGCA_HUMAN
14	155.5	13.9	2333	1	PGCA_CANFA
15	154.5	13.9	612	1	LEM2_MOUSE
16	153.5	13.8	644	1	CD93_MOUSE
17	153.5	13.8	2124	1	PGCA_RAT
18	152	13.6	912	1	PGCB_BOVIN
19	151.5	13.6	652	1	CD93_HUMAN
20	151	13.5	197	1	CLF1_HUMAN
21	149	13.4	321	1	FCE2_HUMAN
22	148.5	13.3	2132	1	PGCA_MOUSE
23	147	13.2	883	1	PGCB_MOUSE
24	146	13.1	2109	1	PGCA_CHICK
25	145	13.0	158	1	LECG_TRIST
26	145	13.0	883	1	PGCB_RAT
27	144.5	13.0	173	1	LEC2_MEGRO
28	144.5	13.0	372	1	LEM1_RAT
29	143.5	12.9	372	1	LEM1_MOUSE
30	141.5	12.7	331	1	FCE2_MOUSE
31	141.5	12.7	549	1	LEM2_RAT
32	140.5	12.6	162	1	LEC3_MEGRO
33	140	12.6	370	1	LEM1_BOVIN

34	139	12.5	248	1	PSPA_HUMAN	P07714	homo sapien
35	139	12.5	283	1	LECA_SARPE	P05047	sarcophaga
36	138.5	12.4	152	1	IXA_TRIFL	P23806	trimeresuru
37	138.5	12.4	202	1	TETN_MOUSE	P43025	mus musculus
38	137.5	12.3	372	1	LEM1_MACMU	Q95198	macaca mula
39	137.5	12.3	372	1	LEM1_PAPHA	Q28768	papio hamad
40	136	12.2	175	1	LITH_BOVIN	P23132	bos taurus
41	135	12.1	165	1	LIT1_MOUSE	P43137	mus musculus
42	134.5	12.1	132	1	ACAL_ANSPY	P83300	anser anser
43	134.5	12.1	372	1	LEM1_PONPY	Q95235	pongo pygma
44	134.5	12.1	485	1	LEM2_BOVIN	P98107	bos taurus
45	134	12.0	166	1	TETN_CARSP	P26258	carcharhinu

ALIGNMENTS

RESULT 1
CHOD_MOUSE
ID CHOD_MOUSE STANDARD; PRT; 273 AA.
AC Q9CXM0; Q8VI31;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Chondrolectin precursor (Transmembrane protein MT75).
GN CHODL.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J;
RA Weng L., Smits P., Hubner R., Wouters J., Merregaert J.;
RT "Mt75, a low expressed c-type lectin gene involving in
chondrogenesis";
RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Embryonic head;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaide I., Pesole G., Quackenbush J.,
RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Momoto N.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weltz C., Whittaker C., Wilming L.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
CC -!- SUBCELLULAR LOCATION: Type I membrane protein (Potential).
CC -!- SIMILARITY: CONTAINS 1 C-TYPE LECTIN FAMILY DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF311699; AAL50354.1; -.

DR EMBL; AK014255; BAB29226.1; -
DR HSSP; P22897; 1EGG.
DR MGD; MGI:1920461; 3110074E07Rik.
DR InterPro; IPR001304; Lectin_C.
DR Pfam; PF00059; lectin_c; 1.
DR SMART; SM00034; CLECT; 1.
DR PROSITE; PS00615; C_TYPE_LLECTIN_1; FALSE_NEG.
DR PROSITE; PS50041; C_TYPE_LLECTIN_2; 1.
KW Lectin; Transmembrane; Glycoprotein; Signal.
FT SIGNAL 1 21 POTENTIAL.
FT CHAIN 22 273 CHONDROLECTIN.
FT DOMAIN 22 216 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 217 237 POTENTIAL.
FT DOMAIN 238 273 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 35 179 C-TYPE LECTIN.
FT CARBOHYD 86 86 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CONFLICT 24 24 V -> W (IN REF. 2).
FT CONFLICT 179 179 T -> K (IN REF. 2).
SQ SEQUENCE 273 AA; 30303 MW; E052D933F244F4C7 CRC64;

Query Match 51.5%; Score 574; DB 1; Length 273;
Best Local Similarity 57.4%; Pred. No. 4e-44;
Matches 109; Conservative 26; Mismatches 41; Indels 14; Gaps 5;

QY 4 RLLSGQPVCRGGTQPCVKVIYFHDTSRRLNFEAEKACRRDGGQLVSIASEDEQKLIK 63
Db :||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
23 RVVSGQKVCFADVKHPCVKMAYFHELSRVFQEARLACESEGGVLLSLENAEQKLIKES 82
QY 64 FIENLLP-----SDGDFWIGLRRREKQSNSTACQDLYAWTDGSIQFRNWWYDEPSCGS 118
Db :||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
83 MLQNLTKPGTIGSDGFWIGLRRSGDGT-SGACPDLYQWSDGSSQFRNWWYDEPSCGS 141
QY 119 EVCVVMYHQPAPAGIGGPFYQWDDRCNMKNFFICKYSDE-KPAVPSRAEAGEETELT 177
Db :||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
142 EKCVMYHQTANPGLGPGYLYQWDDRCNMKNHYICTYPEIHPTEPA-----EKPYL 196
QY 178 TPVLPETOE 187
Db :|||||
197 NQ--PEETHE 204

RESULT 2
CHOD_HUMAN STANDARD; PRT; 273 AA.
AC Q9H9P2; Q9HCY3;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Chondrolectin precursor (Transmembrane protein MT75) (PREDI2
DE protein).
GN CHODL OR C21ORF68.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A., AND TISSUE SPECIFICITY.
RX PubMed=12079284;
RA Weng L., Smits P., Wauters J., Merregaert J.;
RT "Molecular cloning and characterization of human chondrolectin, a
RL novel type I transmembrane protein homologous to C-type lectins.";
RN Genomics 80:62-70(2002).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=20289799; PubMed=10830953;
RA Hattori M., Fujiyama A., Taylor T.D., Watanabe H., Yada T.,
RA Park H.-S., Toyoda A., Ishii K., Totoki Y., Choi D.-K., Soeda E.,
RA Ohki M., Takagi T., Sakaki Y., Taudien S., Blechschmidt K., Polley A.,
RA Menzel U., Delabar J., Kumpf K., Lehmann R., Patterson D.,
RA Reichwald K., Rump A., Schillhabel M., Schudy A., Zimmermann W.,
RA Rosenthal A., Kudoh J., Shibuya K., Kawasaki K., Asakawa S.,
RA Shintani A., Sasaki T., Nagamine K., Mitsuyama S., Antonarakis S.E.,
RA Minoshima S., Shimizu N., Nordsiek G., Hornischer K., Brandt P.,

RA Scharfe M., Schoen O., Desario A., Reichelt J., Kauer G., Bloecker H.,
RA Ramser J., Beck A., Klages S., Hennig S., Rieselmann L., Dagand E.,
RA Wehrmeyer S., Borzym K., Gardiner K., Nizetic D., Francis F.,
RA Lehrach H., Reinhardt R., Yaspo M.-L.;
RT "The DNA sequence of human chromosome 21.";
RL Nature 405:311-319(2000).
RN [3]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RA Strausberg R.;
RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE OF 27-273 FROM N.A.
RA ISOGAI T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
RA Nishikawa T., Nagai K., Sugano S., Shiratori A., Sudo H.,
RA Wagatsuma M., Hosoiri T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,
RA Takahashi M., Chiba Y., Ishida S., Murakawa K., Ono Y., Takiguchi S.,
RA Watanabe S., Kimura K., Murakami K., Ishii S., Kawai Y., Saito K.,
RA Yamamoto J., Wakamatsu A., Nakamura Y., Nagahari K., Masuho Y.,
RA Ninomiya K., Iwayanagi T.;
RT "NEDO human cDNA sequencing project.";
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
RN [5]
RP TISSUE SPECIFICITY.
RX MEDLINE=21564202; PubMed=11707072;
RA Raymond A., Friedli M., Neergard Henrichsen C., Chapot F., Deutsch S.,
RA Ucla C., Rossier C., Lyle R., Guipponi M., Antonarakis S.E.;
RT "From PREDs and open reading frames to cDNA isolation: revisiting the
RL human chromosome 21 transcription map.";
CC Genomics 78:46-54(2001).
CC -|- SUBCELLULAR LOCATION: Type I membrane protein (Potential).
CC -|- TISSUE SPECIFICITY: Found in spleen, testis, prostate and fetal
CC liver. Expression limited to vascular muscle of testis, smooth
CC muscle of prostate stroma, heart muscle, skeletal muscle, crypts
CC of small intestine, and red pulp of spleen.
CC -|- PTM: N-glycosylated.
CC -|- SIMILARITY: CONTAINS 1 C-TYPE LECTIN FAMILY DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/)
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AF257472; AAL05981.1; -
CC EMBL; AL163217; CAB90388.1; -
CC EMBL; BC009418; AAH09418.1; -
CC EMBL; AK022689; BAB14181.1; ALT_INIT.
CC HSSP; P22897; 1EGG.
CC Genew; HGNC:17807; CHODL.
CC InterPro; IPR001304; Lectin_C.
CC Pfam; PF00059; lectin_c; 2.
CC SMART; SM00034; CLECT; 1.
CC PROSITE; PS00615; C_TYPE_LLECTIN_1; FALSE_NEG.
CC PROSITE; PS50041; C_TYPE_LLECTIN_2; 1.
KW Lectin; Transmembrane; Glycoprotein; Signal.
FT SIGNAL 1 21 POTENTIAL.
FT CHAIN 22 273 CHONDROLECTIN.
FT DOMAIN 22 216 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 217 237 POTENTIAL.
FT DOMAIN 238 273 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 35 179 C-TYPE LECTIN.
FT CARBOHYD 86 86 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 273 AA; 30431 MW; F4890AAFB572A311 CRC64;

Query Match 50.4%; Score 561.5; DB 1; Length 273;
Best Local Similarity 60.1%; Pred. No. 5.2e-43;
Matches 101; Conservative 25; Mismatches 35; Indels 7; Gaps 3;

QY 4 RLLSGQPVCRGGTQPCVKVIYFHDTSRRLNFEAEKACRRDGGQLVSIASEDEQKLIK 63
Db :||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||

Db 23 RVVSGQKVCFAFKHPCYKMAYPHELSSRVSFQEARLACESEGGVLLSLENEAEQKLLIES 82

QY 64 FIENLLP-----SDGDFWIGLRREEKQSNSTACQDLYAWTDGSIQFRNWWYDEPSCGS 118

Db 83 MLQNLTKPGTGISDGFWIGLWRNGDGQT-SGACPDLYQWSDGNSQYRNWYTDEPSCGS 141

QY 119 EVCVVMYHQPSAPAGIGGPPYFMQWDDRCNMKNFFICKYSDE-KPAVP 165

Db 142 EKCVMYHQPTANPLGGLPYLYQWDDRCNMKNHYICKYEPEINPTAP 189

RESULT 3

MANR_HUMAN STANDARD; PRT; 1456 AA.

AC P22897;

DT 01-AUG-1991 (Rel. 19, Created)

DT 01-AUG-1991 (Rel. 19, Last sequence update)

DT 15-JUN-2002 (Rel. 41, Last annotation update)

DE Macrophage mannose receptor precursor (MMR) (CD206 antigen).

GN MRC1.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCBI_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.

RC TISSUE=Placenta;

RX MEDLINE=90324192; PubMed=2373685;

RA Taylor M.E., Conary J.T., Lennartz M.R., Stahl P.D., Drickamer K.;

RT "Primary structure of the mannose receptor contains multiple motifs

RT resembling carbohydrate-recognition domains.";

RL J. Biol. Chem. 265:12156-12162(1990).

RN [2]

RP SEQUENCE FROM N.A.

RX MEDLINE=93052405; PubMed=1294118;

RA Kim S.-J., Ruiz N., Bezouska K., Drickamer K.;

RT "Organization of the gene encoding the human macrophage mannose

RT receptor (MRC1).";

RL Genomics 14:721-727(1992).

RN [3]

RP STUDIES ON THE BINDING OF INDIVIDUAL LECTIN DOMAINS.

RX MEDLINE=92112893; PubMed=1730714;

RA Taylor M.E., Bezouska K., Drickamer K.;

RT "Contribution to ligand binding by multiple carbohydrate-recognition

RT domains in the macrophage mannose receptor.";

RL J. Biol. Chem. 267:1719-1726(1992).

RN [4]

RP X-RAY CRYSTALLOGRAPHY (2.3 ANGSTROMS) OF 642-788.

RX MEDLINE=20347275; PubMed=10779515;

RA Feinberg H., Park-Snyder S., Kolatkar A.R., Heise C.T., Taylor M.E.,

RA Weis W.I.;

RT "Structure of a C-type carbohydrate recognition domain from the

RT macrophage mannose receptor.";

RL J. Biol. Chem. 275:21539-21548(2000).

CC -1- FUNCTION: MEDIATES THE ENDOCYTOSIS OF GLYCOPROTEINS BY

CC MACROPHAGES, IN SEVERAL RECOGNITION AND UPTAKE PROCESSES.

CC -1- SUBCELLULAR LOCATION: Type I membrane protein.

CC -1- MISCELLANEOUS: CRDS 1-3 HAVE AT MOST VERY WEAK AFFINITY FOR

CC CARBOHYDRATE. CRD 4 SHOWS THE HIGHEST AFFINITY BINDING AND HAS

CC MULTISPECIFICITY FOR A VARIETY OF MONOSACCHARIDES. AT LEAST 3 CRDS

CC (4, 5, AND 7) ARE REQUIRED FOR HIGH AFFINITY BINDING AND

CC ENDOCYTOSIS OF MULTIVALENT GLYCOCONJUGATES.

CC -1- SIMILARITY: CONTAINS 8 C-TYPE LECTIN FAMILY DOMAINS.

CC -1- SIMILARITY: CONTAINS 1 RICIN B-TYPE LECTIN DOMAIN.

CC -1- DATABASE: NAME=PROW; NOTE=PROW 2:85-89(2001);

CC WWW="http://www.ncbi.nlm.nih.gov/prow/guide/1644341535.g.htm".

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>)

CC or send an email to license@lsb-sib.ch).

CC -----

DR EMBL; J05550; AAA59868.1; -.

DR EMBL; M93221; AAA60389.1; -.

DR EMBL; M93192; AAA60389.1; JOINED.

DR EMBL; M93193; AAA60389.1; JOINED.

DR EMBL; M93194; AAA60389.1; JOINED.

DR EMBL; M93195; AAA60389.1; JOINED.

DR EMBL; M93196; AAA60389.1; JOINED.

DR EMBL; M93197; AAA60389.1; JOINED.

DR EMBL; M93198; AAA60389.1; JOINED.

DR EMBL; M93199; AAA60389.1; JOINED.

DR EMBL; M93200; AAA60389.1; JOINED.

DR EMBL; M93201; AAA60389.1; JOINED.

DR EMBL; M93202; AAA60389.1; JOINED.

DR EMBL; M93203; AAA60389.1; JOINED.

DR EMBL; M93204; AAA60389.1; JOINED.

DR EMBL; M93205; AAA60389.1; JOINED.

DR EMBL; M93206; AAA60389.1; JOINED.

DR EMBL; M93207; AAA60389.1; JOINED.

DR EMBL; M93208; AAA60389.1; JOINED.

DR EMBL; M93209; AAA60389.1; JOINED.

DR EMBL; M93210; AAA60389.1; JOINED.

DR EMBL; M93211; AAA60389.1; JOINED.

DR EMBL; M93212; AAA60389.1; JOINED.

DR EMBL; M93213; AAA60389.1; JOINED.

DR EMBL; M93214; AAA60389.1; JOINED.

DR EMBL; M93215; AAA60389.1; JOINED.

DR EMBL; M93216; AAA60389.1; JOINED.

DR EMBL; M93217; AAA60389.1; JOINED.

DR EMBL; M93218; AAA60389.1; JOINED.

DR EMBL; M93219; AAA60389.1; JOINED.

DR EMBL; M93220; AAA60389.1; JOINED.

DR PIR; A36563; A36563.

DR PDB; 1EGG; 30-AUG-00.

DR PDB; 1EGI; 30-AUG-00.

DR Genew; HGNC:7228; MRC1.

DR MIM; 153618; -.

DR InterPro; IPR000562; FN_Type_II.

DR InterPro; IPR001304; Lectin_C.

DR InterPro; IPR000772; Ricin_B_lectin.

DR Pfam; PF00040; fn2; 1.

DR Pfam; PF00059; lectin_c; 8.

DR Pfam; PF00652; Ricin_B_lectin; 2.

DR PRINTS; PR00013; FNTYPEII.

DR ProDom; PD000995; FN_Type_II; 1.

DR SMART; SM00034; CLECT; 8.

DR SMART; SM00059; FN2; 1.

DR SMART; SM00458; RICIN; 1.

DR PROSITE; PS00615; C_TYPE_LLECTIN_1; 6.

DR PROSITE; PS50041; C_TYPE_LLECTIN_2; 8.

DR PROSITE; PS00023; FIBRONECTIN_2; 1.

DR PROSITE; PS50231; RICIN_B_LLECTIN; 1.

KW Receptor; Signal; Calcium-binding; Transmembrane; Repeat;

KW Glycoprotein; Endocytosis; Macrophage; Lectin; Antigen; 3D-structure.

FT SIGNAL 1 18

FT CHAIN 19 1456

FT DOMAIN 19 1383

FT TRANSMEM 1384 1411

FT DOMAIN 1412 1456

FT DOMAIN 22 142

FT DOMAIN 157 212

FT DOMAIN 216 344

FT DOMAIN 360 490

FT DOMAIN 502 629

FT DOMAIN 644 781

FT DOMAIN 805 926

FT DOMAIN 943 1083

FT DOMAIN 1100 1216

FT DOMAIN 1228 1359

FT DISULFID 646 659

FT DISULFID 680 777

FT DISULFID 753

FT MACROPHAGE MANNOSE RECEPTOR.

FT EXTRACELLULAR (POTENTIAL).

FT POTENTIAL.

FT CYTOPLASMIC (POTENTIAL).

FT RICIN B-TYPE LECTIN.

FT FIBRONECTIN TYPE-II.

FT C-TYPE LECTIN 1 (LONG FORM).

FT C-TYPE LECTIN 2 (LONG FORM).

FT C-TYPE LECTIN 3 (LONG FORM).

FT C-TYPE LECTIN 4 (LONG FORM).

FT C-TYPE LECTIN 5 (LONG FORM).

FT C-TYPE LECTIN 6 (LONG FORM).

FT C-TYPE LECTIN 7 (LONG FORM).

FT C-TYPE LECTIN 8 (LONG FORM).

OY 134 IGGPYMFOWNDRCNMKNFICK 156
Db 1146 -----RWNDVPCNYNLPYVCK 1161

RESULT 5
PGCV_BOVIN STANDARD; PRT; 3381 AA.
AC P81282; 077609; 077610; 077611; 077612;
DT 15-DEC-1998 (Rel. 37, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Versican core protein precursor (Large fibroblast proteoglycan)
DE (Chondroitin sulfate proteoglycan core protein 2) (PG-M) (Glial
DE hyaluronate-binding protein) (GHAP).
GN CP5G2.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORMS V0; V1; V2 AND V3).
RC TISSUE=Forebrain;
RX MEDLINE-9828320; PubMed-9624174;
RA Schmalfeldt M., Dours-Zimmermann M.T., Winterhalter K.H.,
RA Zimmermann D.R.;
RT "Versican V2 is a major extracellular matrix component of the mature
RT bovine brain."
RL J. Biol. Chem. 273:15758-15764(1998).
RN [2]
RP SEQUENCE OF 21-53; 78-96; 226-250; 262-277; 295-306; 314-324; 329-331
RP AND 342-348.
RX TISSUE-Spinal cord;
RC MEDLINE-92062692; PubMed-1720020;
RA Perides G., Biviano F., Bignami A.;
RT "Interaction of a brain extracellular matrix protein with hyaluronic
RT acid."
RL Biochim. Biophys. Acta 1075:248-258(1991).
CC -!- FUNCTION: May play a role in intercellular signaling and in
CC connecting cells with the extracellular matrix. May take part in
CC the regulation of cell motility, growth and differentiation. Binds
CC hyaluronic acid.
CC -!- SUBCELLULAR LOCATION: Secreted; extracellular matrix.
CC -!- ALTERNATIVE PRODUCTS: At least 4 isoforms; V0 (shown here), V1, V2
CC and V3; are produced by alternative splicing.
CC -!- TISSUE SPECIFICITY: Cerebral white matter. V0 and V1 are expressed
CC in the central nervous system, and in a number of mesenchymal and
CC epithelial tissues; the major isoform v2 is restricted to the
CC central nervous system.
CC -!- DEVELOPMENTAL STAGE: Disappears after the cartilage development
CC (by similarity).
CC -!- SIMILARITY: CONTAINS 1 IMMUNOGLOBULIN-LIKE V-TYPE DOMAIN.
CC -!- SIMILARITY: CONTAINS 2 LINK DOMAINS.
CC -!- SIMILARITY: CONTAINS 2 EGF-LIKE DOMAINS.
CC -!- SIMILARITY: CONTAINS 1 C-TYPE LECTIN FAMILY DOMAIN.
CC -!- SIMILARITY: CONTAINS 1 SUSHI (SCR) DOMAIN.
CC -!- SIMILARITY: BELONGS TO THE AGGRECAN/VERSICAN PROTEOGLYCAN FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL; AF060456; AAC24358.1; -
DR EMBL; AF060457; AAC24359.1; -
DR EMBL; AF060458; AAC24360.1; -
DR EMBL; AF060459; AAC24361.1; -
DR HSSP; P01132; IEPG.
DR InterPro; IPR000152; Asx_hydroxyl.

DR InterPro; IPR000561; EGF-like.
DR InterPro; IPR000742; EGF_2.
DR InterPro; IPR001881; EGF_Ca.
DR InterPro; IPR003599; Ig.
DR InterPro; IPR003006; Ig_MHC.
DR InterPro; IPR003600; Ig_like.
DR InterPro; IPR001304; Lectin_C.
DR InterPro; IPR000538; Link.
DR InterPro; IPR000436; Sushi_SCR_CCP.
DR Pfam; PF00008; EGF; 2.
DR Pfam; PF00047; Ig; 1.
DR Pfam; PF00059; lectin_c; 1.
DR Pfam; PF00084; sushi; 1.
DR Pfam; PF00193; Xlink; 2.
DR PRINTS; PR01265; LINKMODULE.
DR ProDom; PD000918; Link; 1.
DR SMART; SM00032; CCP; 1.
DR SMART; SM00034; CLECT; 1.
DR SMART; SM00181; EGF; 2.
DR SMART; SM00179; EGF_Ca; 2.
DR SMART; SM00409; IG; 1.
DR SMART; SM00410; IG_like; 1.
DR SMART; SM00445; LINK; 2.
DR PROSITE; PS00010; ASX_HYDROXYL; 1.
DR PROSITE; PS00615; C_TYPE_LECTIN_1; 1.
DR PROSITE; PS50041; C_TYPE_LECTIN_2; 1.
DR PROSITE; PS00022; EGF_1; 2.
DR PROSITE; PS01186; EGF_2; 1.
DR PROSITE; PS01187; EGF_Ca; 1.
DR PROSITE; PS01241; LINK; 2.
KW Glycoprotein; Proteoglycan; Lectin; Extracellular matrix; Sushi;
KW Signal; Repeat; EGF-like domain; Calcium; Immunoglobulin domain;
KW Hyaluronic acid; Alternative splicing.
FT SIGNAL 1 20 POTENTIAL.
FT CHAIN 21 3381 VERSICAN CORE PROTEIN.
FT DOMAIN 37 138 IG-LIKE V-TYPE DOMAIN.
FT DOMAIN 168 245 LINK 1.
FT DOMAIN 266 347 LINK 2.
FT DOMAIN 349 1336 GAG-ALPHA (GLUCOSAMINOGLYCAN ATTACHMENT
FT DOMAIN 1337 3074 DOMAIN).
FT DOMAIN 3074 3110 GAG-BETA.
FT DOMAIN 3112 3148 EGF-LIKE 1.
FT DOMAIN 3161 3275 EGF-LIKE 2, CALCIUM-BINDING.
FT DOMAIN 3280 3338 C-TYPE LECTIN.
FT DISULFID 44 131 SUSHI.
FT DISULFID 173 244 BY SIMILARITY.
FT DISULFID 197 218 BY SIMILARITY.
FT DISULFID 271 346 BY SIMILARITY.
FT DISULFID 295 316 BY SIMILARITY.
FT DISULFID 3078 3089 BY SIMILARITY.
FT DISULFID 3083 3098 BY SIMILARITY.
FT DISULFID 3100 3109 BY SIMILARITY.
FT DISULFID 3116 3127 BY SIMILARITY.
FT DISULFID 3121 3136 BY SIMILARITY.
FT DISULFID 3138 3147 BY SIMILARITY.
FT DISULFID 3154 3165 BY SIMILARITY.
FT DISULFID 3182 3274 BY SIMILARITY.
FT DISULFID 3250 3266 BY SIMILARITY.
FT DISULFID 3281 3324 BY SIMILARITY.
FT DISULFID 3310 3337 BY SIMILARITY.
FT CARBOHYD 57 57 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 331 331 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 352 352 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 817 817 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 965 965 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1017 1017 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1333 1333 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1393 1393 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1437 1437 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1463 1463 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1653 1653 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1974 1974 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT	CARBOHYD	2045	2045	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	2074	2074	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	2103	2103	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	2263	2263	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	2290	2290	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	2356	2356	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	2623	2623	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	2641	2641	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	2919	2919	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	3052	3052	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	3354	3354	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	3364	3364	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	VARSP LIC	349	349	P -> R (IN ISOFORM V1 AND V3).
FT	VARSP LIC	350	1336	MISSING (IN ISOFORM V1).
FT	VARSP LIC	1337	3074	MISSING (IN ISOFORM V2).
FT	VARSP LIC	350	3074	MISSING (IN ISOFORM V3).
FT	CONFLICT	25	25	MISSING (IN REF. 2).
FT	CONFLICT	51	51	MISSING (IN REF. 2).
FT	CONFLICT	89	89	N -> D (IN REF. 2).
FT	CONFLICT	96	96	Q -> D (IN REF. 2).
FT	CONFLICT	346	346	C -> R (IN REF. 2).
SQ	SEQUENCE	3381 AA;	369984 MW;	F09716FA7778D459 CRC64;
Query Match 15.7%; Score 174.5; DB 1; Length 3381;				
Best Local Similarity 25.5%; Pred. No. 3.5e-07;				
Matches 59; Conservative 26; Mismatches 71; Indels 75; Gaps 11;				
QY	17	QRPCYKVIYFHDTSRLNFEEAKEACRRDGGQLVSI ESEDEQKLIKFIENLLPSDGDGDF- 75		
Db	3162	QGQCYK--YF---AHRRTWDAARECRLOGAHLTSLSHEEQMFVNRV-----GHDYQ 3209		
QY	76	WIGLRRREKQSNSTACQDLYAWTDGSI SQFNWYVDEP----SCGSEVCVVMYHQPSAP 131		
Db	3210	WIGL-----NDKMFEDHFRWTDGSLQYENWRPNQDPSFFSTGDCVVIWHENG-- 3259		
QY	132	AGIGGPMYFQWDDRCNMKNFICKYS-----DEKPAVPSRAEAGE----- 172		
Db	3260	-----QWNDVPCNYHLTYTCKKGTVACGPPVVENAKTFGKMKPRYEINSLIRYHC 3310		
QY	173	-----ELETLT-----PVL-----PEETQEEDAKKTFKESREAA LN 203		
Db	3311	KDGFQIRHLPTIRCLNGRWPAMPKITCLNPSAYQRTYSKKYFNKSSSAKD N 3361		
RESULT 6				
PGCN_RAT		STANDARD;	PRT;	1257 AA.
ID	PGCN_RAT			
AC	P55067;			
DT	01-OCT-1996 (Rel. 34, Created)			
DT	01-OCT-1996 (Rel. 34, Last sequence update)			
DT	16-OCT-2001 (Rel. 40, Last annotation update)			
DE	Neurocan core protein precursor (245 kDa early postnatal core glycoprotein) [Contains: 150 kDa adult core glycoprotein].			
GN	CSPG3 OR NCAN.			
OS	Rattus norvegicus (Rat).			
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.			
OX	NCBI_TaxID=10116;			
RN	[1]			
RP	SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.			
RC	STRAIN-Sprague-Dawley; TISSUE=Brain;			
RX	MEDLINE=92406907; PubMed=1326557;			
RA	Rauch U., Karthikeyan L., Maurel P., Margolis R.U., Margolis R.K.;			
RT	"Cloning and primary structure of neurocan, a developmentally regulated, aggregating chondroitin sulfate proteoglycan of brain.";			
RL	J. Biol. Chem. 267:19536-19547(1992).			
RN	[2]			
RP	CHARACTERIZATION.			
RX	MEDLINE=94230574; PubMed=7513709;			
RA	Friedlander D.R., Milev P., Karthikeyan L., Margolis R.K.,			
RA	Margolis R.U., Grumet M.;			
RT	"The neuronal chondroitin sulfate proteoglycan neurocan binds to the neural cell adhesion molecules Ng-CAM/L1/NILE and N-CAM, and inhibits			

RT	neuronal adhesion and neurite outgrowth.*;
RL	J. Cell Biol. 125:669-680(1994).
CC	-1- FUNCTION: MAY MODULATE NEURONAL ADHESION AND NEURITE GROWTH DURING DEVELOPMENT BY BINDING TO NEURAL CELL ADHESION MOLECULES (NG-CAM AND N-CAM). CHONDROITIN SULFATE PROTEOGLYCAN; BINDS TO HYALURONIC ACID.
CC	-1- TISSUE SPECIFICITY: EARLY POSTNATAL AND ADULT BRAIN; NOT EXPRESSED IN KIDNEY, LUNG, LIVER AND MUSCLE.
CC	-1- PTM: CONTAINS MOSTLY CHONDROITIN SULFATE, BUT ALSO N-LINKED AND O-LINKED OLIGOSACCHARIDES (BY SIMILARITY).
CC	-1- PTM: TWO ISOFORMS WERE FOUND THAT PROBABLY ARISE BY PROTEOLYTIC DEGRADATION. THE LARGE ISOFORM IS PREDOMINANT IN EARLY POSTNATAL BRAIN, THE SMALL ISOFORM IS FOUND IN ADULT BRAIN.
CC	-1- SIMILARITY: CONTAINS 1 IMMUNOGLOBULIN-LIKE V-TYPE DOMAIN.
CC	-1- SIMILARITY: CONTAINS 2 EGF-LIKE DOMAINS.
CC	-1- SIMILARITY: CONTAINS 2 LINK DOMAINS.
CC	-1- SIMILARITY: CONTAINS 1 C-TYPE LECTIN FAMILY DOMAIN.
CC	-1- SIMILARITY: CONTAINS 1 SUSHI (SCR) DOMAIN.
CC	-1- SIMILARITY: BELONGS TO THE AGGRECAN/VERSICAN PROTEOGLYCAN FAMILY.
CC	-----
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
CC	-----
DR	EMBL; M97161; AAC37679.1; -.
DR	HSP; P00740; 1EDM.
DR	InterPro; IPR000152; Asx_hydroxyl.
DR	InterPro; IPR000561; EGF-like.
DR	InterPro; IPR000742; EGF_2.
DR	InterPro; IPR001881; EGF_Ca.
DR	InterPro; IPR003599; Ig.
DR	InterPro; IPR003006; Ig_MHC.
DR	InterPro; IPR001304; Lectin_C.
DR	InterPro; IPR000538; Link.
DR	InterPro; IPR000436; Sushi_SCR_CCP.
DR	Pfam; PF00008; EGF; 2.
DR	Pfam; PF00047; Ig; 1.
DR	Pfam; PF00059; lectin_c; 1.
DR	Pfam; PF00084; sushi; 1.
DR	Pfam; PF00193; Xlink; 2.
DR	ProDom; PD000918; Link; 2.
DR	SMART; SM00032; CCP; 1.
DR	SMART; SM00034; CLECT; 1.
DR	SMART; SM00179; EGF_CA; 1.
DR	SMART; SM00001; EGF_like; 1.
DR	SMART; SM00409; IG; 1.
DR	SMART; SM00445; LINK; 2.
DR	PROSITE; PS00010; ASX_HYDROXYL; 1.
DR	PROSITE; PS00022; EGF_1; 3.
DR	PROSITE; PS01186; EGF_2; 1.
DR	PROSITE; PS01187; EGF_CA; 1.
DR	PROSITE; PS01241; LINK; 2.
DR	PROSITE; PS00615; C_TYPE_LLECTIN_1; 1.
DR	PROSITE; PS50041; C_TYPE_LLECTIN_2; 1.
KW	Glycoprotein; Hyaluronic acid; Proteoglycan; Immunoglobulin domain;
KW	EGF-like domain; Calcium; Repeat; Lectin; Sushi; Signal.
FT	SIGNAL 1 22
FT	CHAIN 23 1257
FT	CHAIN 639 1257
FT	DOMAIN 51 146
FT	DOMAIN 176 253
FT	DOMAIN 274 355
FT	DOMAIN 949 985
FT	DOMAIN 987 1023
FT	DOMAIN 1025 1154
FT	DOMAIN 1155 1213
FT	DISULFID 58 139
FT	DISULFID 181 252
FT	DISULFID 205 226
	NEUROCAN CORE PROTEIN.
	150 KDA ADULT CORE GLYCOPROTEIN.
	IG-LIKE V-TYPE DOMAIN.
	LINK 1.
	LINK 2.
	EGF-LIKE 1.
	EGF-LIKE 2, CALCIUM-BINDING (POTENTIAL).
	C-TYPE LECTIN.
	SUSHI.
	BY SIMILARITY.
	BY SIMILARITY.
	BY SIMILARITY.


```
FT DISULFID 279 354 BY SIMILARITY.
FT DISULFID 303 324 BY SIMILARITY.
FT DISULFID 953 964 BY SIMILARITY.
FT DISULFID 958 973 BY SIMILARITY.
FT DISULFID 975 984 BY SIMILARITY.
FT DISULFID 1029 1040 BY SIMILARITY.
FT DISULFID 1057 1149 BY SIMILARITY.
FT DISULFID 1125 1141 BY SIMILARITY.
FT DISULFID 1156 1199 BY SIMILARITY.
FT DISULFID 1185 1212 BY SIMILARITY.
FT CARBOHYD 121 121 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 339 339 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 737 737 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 944 944 O-LINKED (XYL. . .) (CHONDROITIN SULFATE).
FT CARBOHYD 967 967 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1164 1164 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 1257 AA; 135544 MW; 992B33DCFA19EE1B CRC64;

Query Match 15.6%; Score 174; DB 1; Length 1257;
Best Local Similarity 30.8%; Pred. No. 1.2e-07;
Matches 44; Conservative 18; Mismatches 49; Indels 32; Gaps 6;

QY 17 QRPCYKVIYFHTSRLNFEAKEACRRDGGQLVSIESEDEQKLEKFIENLLPSDGF 76
Db 1037 QGHCYR--YF---AHRRAWEDAERDCRRRAGHLTSVHSPEHKFINSF-----GHENSW 1085

QY 77 IGLRRREKQSNSTACQDLYAWTDGSIQFRNWMYVDEPS---CGSEVCVMYHQPSAPAG 133
Db 1086 IGLNDRTVRD-----FQWTDNTGLQYENWREKQPDNFFAGGEDCVMVAHENG--- 1134

QY 134 IGGPYMFQWDDRCNMKNFNICK 156
Db 1135 -----RWNDVPCYNLPHYCK 1150

RESULT 7
PGCV_RAT STANDARD; PRT; 2738 AA.
AC Q9ERB4; O08592; O08564; Q9R1K4;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Versican core protein precursor (Large fibroblast proteoglycan)
DE (Chondroitin sulfate proteoglycan core protein 2) (PG-M) (Glial
DE hyaluronate-binding protein) (GHAP) (Fragments).
GN CP5G2.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Euthera; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE OF 349-2738 FROM N.A. (ISOFORM V0), SEQUENCE FROM N.A.
RP (ISOFORM V3), AND SEQUENCE OF 2657-2738 FROM N.A. (ISOFORM VINT).
RC STRAIN=Wistar Kyoto;
RX MEDLINE=99327053; PubMed=10397680;
RA Lemire J.M., Braun K.R., Maurel P., Kaplan E.D., Schwartz S.M.,
RA Wight T.N.;
RT "Versican/PG-M isoforms in vascular smooth muscle cells.";
RL Arterioscler. Thromb. Vasc. Biol. 19:1630-1639(1999).
RN [2]
RP SEQUENCE OF 349-2738 FROM N.A. (ISOFORM V0).
RC STRAIN=Wistar Kyoto;
RX MEDLINE=98308094; PubMed=9642104;
RA Milev P., Maurel P., Chiba A., Mevissen M., Popp S., Yamaguchi Y.,
RA Margolis R.K., Margolis R.U.;
RT "Differential regulation of expression of hyaluronan-binding
RT proteoglycans in developing brain: aggrecan, versican, neurocan, and
RL Biochem. Biophys. Res. Commun. 247:207-212(1998).
RN [3]
RP SEQUENCE OF 2421-2463 FROM N.A. (ISOFORM V0).
RC TISSUE-Kidney;
RX MEDLINE=98094159; PubMed=9434070;
```

```
RA Pyke C., Kristensen P., Ostergaard P.B., Oturai P.S., Romer J.;
RT "Proteoglycan expression in the normal rat kidney.";
RL Nephron 77:461-470(1997).
RN [4]
RP SEQUENCE OF 2535-2738 FROM N.A.
RC STRAIN=Sprague-Dawley; TISSUE=Lung;
RA Blomberg L.A., Chan W.-Y., Clerch L., Massaro D.;
RT "Molecular cloning and characterization of two developmentally
RT regulated genes in rat lung.";
RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: May play a role in intercellular signaling and in
CC connecting cells with the extracellular matrix. May take part in
CC the regulation of cell motility, growth and differentiation. Binds
CC hyaluronic acid.
CC -!- SUBCELLULAR LOCATION: Secreted; extracellular matrix.
CC -!- ALTERNATIVE PRODUCTS: At least 3 isoforms; V0 (shown here), V3 and
CC Vint; are produced by alternative splicing.
CC -!- TISSUE SPECIFICITY: In kidney is expressed in the papillary area,
CC but not in glomeruli.
CC -!- DEVELOPMENTAL STAGE: Disappears after the cartilage development
CC (By similarity).
CC -!- SIMILARITY: CONTAINS 1 IMMUNOGLOBULIN-LIKE V-TYPE DOMAIN.
CC -!- SIMILARITY: CONTAINS 2 LINK DOMAINS.
CC -!- SIMILARITY: CONTAINS 2 EGF-LIKE DOMAINS.
CC -!- SIMILARITY: CONTAINS 1 C-TYPE LECTIN FAMILY DOMAIN.
CC -!- SIMILARITY: CONTAINS 1 SUSHI (SCR) DOMAIN.
CC -!- SIMILARITY: BELONGS TO THE AGGREGAN/VERSICAN PROTEOGLYCAN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF062402; AAC40166.1; -.
DR EMBL; U75306; AAB51125.1; -.
DR EMBL; AF084544; AAD48544.1; -.
DR EMBL; AF072892; AAC26116.1; -.
DR EMBL; AY007691; AAG16631.1; -.
DR HSSP; P01132; 1EPG.
DR InterPro; IPR000152; Asx_hydroxyl.
DR InterPro; IPR000561; EGF-like.
DR InterPro; IPR000742; EGF_2.
DR InterPro; IPR001881; EGF_Ca.
DR InterPro; IPR003599; Ig.
DR InterPro; IPR003600; Ig_like.
DR InterPro; IPR001304; Lectin_C.
DR InterPro; IPR000538; Link.
DR InterPro; IPR000436; Sushi_SCR_CCP.
DR PRINTS; PR01265; LINKMODULE.
DR SMART; SM00032; CCP; 1.
DR SMART; SM00034; CLECT; 1.
DR SMART; SM00181; EGF; 2.
DR SMART; SM00179; EGF_CA; 2.
DR SMART; SM00409; IG; 1.
DR SMART; SM00410; IG_like; 1.
DR SMART; SM00445; LINK; 2.
DR PROSITE; PS00010; ASX_HYDROXYL; 1.
DR PROSITE; PS00615; C_TYPE_LECTIN_1; 1.
DR PROSITE; PS50041; C_TYPE_LECTIN_2; 1.
DR PROSITE; PS00022; EGF_1; 2.
DR PROSITE; PS01186; EGF_2; 1.
DR PROSITE; PS01187; EGF_CA; 1.
DR PROSITE; PS01241; LINK; 2.
KW Glycoprotein; Proteoglycan; Lectin; Extracellular matrix; Sushi;
KW Signal; Repeat; EGF-like domain; Calcium; Immunoglobulin domain;
KW Hyaluronic acid; Alternative splicing.
FT SIGNAL 1 20 POTENTIAL.
FT CHAIN 21 2738 VERSICAN CORE PROTEIN.
FT NON_CONS 348 349
FT DOMAIN 37 137 IG-LIKE V-TYPE DOMAIN.
```

FT	DOMAIN	167	244	LINK 1.	
FT	DOMAIN	265	346	LINK 2.	
FT	DOMAIN	<349	695	GAG-ALPHA (GLUCOSAMINOGLYCAN ATTACHMENT DOMAIN).	
FT	DOMAIN			GAG-BETA.	
FT	DOMAIN	696	2431	EGF-LIKE 1.	
FT	DOMAIN	2431	2467	EGF-LIKE 2.	
FT	DOMAIN	2469	2505	C-TYPE LECTIN.	
FT	DOMAIN	2518	2632	SUSHI.	
FT	DOMAIN	2637	2695	BY SIMILARITY.	
FT	DISULFID	44	130	BY SIMILARITY.	
FT	DISULFID	172	243	BY SIMILARITY.	
FT	DISULFID	196	217	BY SIMILARITY.	
FT	DISULFID	270	345	BY SIMILARITY.	
FT	DISULFID	294	315	BY SIMILARITY.	
FT	DISULFID	2435	2446	BY SIMILARITY.	
FT	DISULFID	2440	2455	BY SIMILARITY.	
FT	DISULFID	2457	2466	BY SIMILARITY.	
FT	DISULFID	2473	2484	BY SIMILARITY.	
FT	DISULFID	2478	2493	BY SIMILARITY.	
FT	DISULFID	2495	2504	BY SIMILARITY.	
FT	DISULFID	2511	2522	BY SIMILARITY.	
FT	DISULFID	2539	2631	BY SIMILARITY.	
FT	DISULFID	2607	2623	BY SIMILARITY.	
FT	DISULFID	2638	2681	BY SIMILARITY.	
FT	DISULFID	2667	2694	BY SIMILARITY.	
FT	CARBOHYD	57	57	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	330	330	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	692	692	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	758	758	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	802	802	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	805	805	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	1257	1257	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	1435	1435	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	1633	1633	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	1660	1660	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	1684	1684	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	1738	1738	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	1848	1848	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	2004	2004	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	2409	2409	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	2711	2711	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	2721	2721	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	VARSP LIC	349	2431	MISSING (IN ISOFORM V3).	
FT	VARSP LIC	2697	2738	PSAYQRTYSKKYLKNSSSVKDNSINTSKHEHRWSRRWQETR	
FT	CONFLICT	2535	2539	R -> RKWSFRKNGQPCFNKY (IN ISOFORM VINT).	
FT	SEQUENCE	2738 AA;	300004 MW;	AEREC -> NSARG (IN REF. 4).	
FT	SEQUENCE	2738 AA;	300004 MW;	12CA626D58BD8C6A-CRC64;	
Query Match					15.6%; Score 174; DB 1; Length 2738;
Best Local Similarity					28.5%; Pred. No. 3e-07;
Matches					47; Conservative 23; Mismatches 55; Indels 40; Gaps 8;
QY	17	QRPCYKVIYFHDTSRRLNFEEAKEACRRDGGOLVSISEDEQKLIKFIENLLPSDGF-	75		
Db	2519	QGQCYK--YF---AHRRTWDAARECRLOGAHLTSILSHEEQMFVNRV-----GHDYQ	2566		
QY	76	WIGLRRREEKQSNSTACQDLXAWTDGSIQFRNYYVDEP---SCGSEVCVVMYHQPSAP	131		
Db	2567	WIGL-----NDKMFEDHDFRTDGSALQYENWRPNQPDFFSAGEDCVVIWHENG--	2616		
QY	132	AGIGGYPYMFQWDDRCNMKNFICKYS-----DEKPAVPSREAEGE	172		
Db	2617	-----QWNDVPCNYHLTYTCKKGTVACGQPPVVENAKTFGK	2652		
RESULT 8					
PGCV_MOUSE					
ID	PGCV_MOUSE	STANDARD;	PRT;	3358	AA.
AC	Q62059; Q62058; Q9CUU0;				
DT	01-NOV-1997 (Rel. 35, Created)				
DT	01-NOV-1997 (Rel. 35, Last sequence update)				
DT	16-OCT-2001 (Rel. 40, Last annotation update)				
DE	Versican core protein precursor (Large fibroblast proteoglycan)				

DE	(Chondroitin sulfate proteoglycan core protein 2) (PG-M).
GN	CSPG2.
OS	Mus musculus (Mouse).
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX	NCBI_TaxID=10090;
RN	[1]
RP	SEQUENCE FROM N.A. (ISOFORMS V0; V1 AND V2).
RC	STRAIN=C57BL/6, and Swiss Webster; TISSUE=Brain;
RX	MEDLINE=95122551; PubMed=7822336;
RA	Ito K., Shinomura T., Zako M., Ujita M., Kimata K.;
RT	"Multiple forms of mouse PG-M, a large chondroitin sulfate
RT	proteoglycan generated by alternative splicing.";
RL	J. Biol. Chem. 270:958-965(1995).
RN	[2]
RP	SEQUENCE FROM N.A. (ISOFORM V3).
RC	STRAIN=C57BL/6;
RX	MEDLINE=95181355; PubMed=7876137;
RA	Zako M., Shinomura T., Ujita M., Ito K., Kimata K.;
RT	"Expression of PG-M(V3), an alternatively spliced form of PG-M
RT	without a chondroitin sulfate attachment in region in mouse and human
RT	tissues.";
RL	J. Biol. Chem. 270:3914-3918(1995).
RN	[3]
RP	SEQUENCE OF 1-1692 FROM N.A. (ISOFORM V1).
RC	STRAIN=C57BL/6J; TISSUE=Skin;
RX	MEDLINE=21085660; PubMed=11217851;
RA	Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA	Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA	Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA	Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA	Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA	Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA	Kuehl P., Lewis S., Matsuo Y., Nikaide I., Pesole G., Quackenbush J.,
RA	Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA	Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA	Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA	Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA	Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA	Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA	Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA	Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA	Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA	Wyshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA	Hayashizaki Y.;
RA	"Functional annotation of a full-length mouse cDNA collection.";
RT	Nature 409:685-690(2001).
RL	
CC	-!- FUNCTION: May play a role in intercellular signaling and in
CC	connecting cells with the extracellular matrix. May take part in
CC	the regulation of cell motility, growth and differentiation. Binds
CC	hyaluronic acid.
CC	-!- SUBCELLULAR LOCATION: Secreted; extracellular matrix.
CC	-!- ALTERNATIVE PRODUCTS: At least 4 isoforms; V0 (shown here), V1,
CC	V2 and V3; are produced by alternative splicing.
CC	-!- TISSUE SPECIFICITY: V2 is found only in brain.
CC	-!- DEVELOPMENTAL STAGE: Disappears after the cartilage development.
CC	-!- SIMILARITY: CONTAINS 1 IMMUNOGLOBULIN-LIKE V-TYPE DOMAIN.
CC	-!- SIMILARITY: CONTAINS 2 LINK DOMAINS.
CC	-!- SIMILARITY: CONTAINS 2 EGF-LIKE DOMAINS.
CC	-!- SIMILARITY: CONTAINS 1 C-TYPE LECTIN FAMILY DOMAIN.
CC	-!- SIMILARITY: CONTAINS 1 SUSHI (SCR) DOMAIN.
CC	-!- SIMILARITY: BELONGS TO THE AGGREGAN/VERSICAN PROTEOGLYCAN FAMILY.
CC	-----
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC	the European Bioinformatics Institute. There are no restrictions on its
CC	use by non-profit institutions as long as its content is in no way
CC	modified and this statement is not removed. Usage by and for commercial
CC	entities requires a license agreement (See http://www.isb-sib.ch/
CC	or send an email to license@isb-sib.ch).
CC	-----
DR	EMBL; D16263; BAA03796.1; -.
DR	EMBL; D28599; -; NOT_ANNOTATED_CDS.

[4] SEQUENCE OF 2711-3396 FROM N.A.
RC TISSUE=Lung fibroblast;
RX MEDLINE-88007514; PubMed-2820964;
RA Krusius T., Gehlsen K.R., Ruoslahti E.;
RT "A fibroblast chondroitin sulfate proteoglycan core protein contains
RL lectin-like and growth factor-like sequences.";
RN J. Biol. Chem. 262:13120-13125(1987).
[5]
RP SEQUENCE OF 251-347 FROM N.A.
RX MEDLINE-93122792; PubMed-1478664;
RA Iozzo R.V., Naso M.F., Cannizzaro L.A., Wasmuth J.J.,
RA McPherson J.D.;
RT "Mapping of the versican proteoglycan gene (CSPG2) to the long arm of
RL human chromosome 5 (5q12-5q14).";
RN Genomics 14:845-851(1992).
[6]
RP SEQUENCE FROM N.A. (ISOFORM V3).
RC TISSUE=Brain;
RX MEDLINE-95181355; PubMed-7876137;
RA Zako M., Shinomura T., Ujita M., Ito K., Kimata K.;
RT "Expression of PG-M(V3), an alternatively spliced form of PG-M
RL without a chondroitin sulfate attachment in region in mouse and human
RN tissues.";
RP J. Biol. Chem. 270:3914-3918(1995).
[7]
RP SEQUENCE OF 3333-3396 FROM N.A. (ISOFORM VINT).
RC TISSUE=Aortic smooth muscle;
RX MEDLINE-99327053; PubMed-10397680;
RA Lemire J.M., Braun K.R., Maurel P., Kaplan E.D., Schwartz S.M.,
RA Wight T.N.;
RT "Versican/PG-M isoforms in vascular smooth muscle cells.";
RL Arterioscler. Thromb. Vasc. Biol. 19:1630-1639(1999).
[8]
RP PARTIAL SEQUENCE.
RC TISSUE=Brain;
RX MEDLINE-89174663; PubMed-2466833;
RA Perides G., Lane W.S., Andrews D., Dahl D., Bignami A.;
RT "Isolation and partial characterization of a glial
RL hyaluronate-binding protein.";
RN J. Biol. Chem. 264:5981-5987(1989).
[9]
RP TISSUE SPECIFICITY OF ISOFORMS.
RX MEDLINE-96213482; PubMed-8627343;
RA Paulus W., Baur I., Dours-Zimmermann M.T., Zimmermann D.R.;
RT "Differential expression of versican isoforms in brain tumors.";
RL J. Neuropathol. Exp. Neurol. 55:528-533(1996).
CC -1- FUNCTION: May play a role in intercellular signaling and in
CC connecting cells with the extracellular matrix. May take part in
CC the regulation of cell motility, growth and differentiation. Binds
CC hyaluronic acid.
CC -1- SUBCELLULAR LOCATION: Secreted; extracellular matrix.
CC -1- ALTERNATIVE PRODUCTS: At least 5 isoforms; V0 (shown here), V1,
CC V2, V3 and Vint; are produced by alternative splicing.
CC -1- TISSUE SPECIFICITY: Cerebral white matter. V0 and V1 is expressed
CC in normal brain, gliomas, medulloblastomas, schwannomas,
CC neurofibromas, and meningiomas; v2 is restricted to normal brain
CC and gliomas; v3 is found in all these tissues except
CC medulloblastomas.
CC -1- DEVELOPMENTAL STAGE: Disappears after the cartilage development.
CC -1- SIMILARITY: CONTAINS 1 IMMUNOGLOBULIN-LIKE V-TYPE DOMAIN.
CC -1- SIMILARITY: CONTAINS 2 LINK DOMAINS.
CC -1- SIMILARITY: CONTAINS 2 EGF-LIKE DOMAINS.
CC -1- SIMILARITY: CONTAINS 1 C-TYPE LECTIN FAMILY DOMAIN.
CC -1- SIMILARITY: CONTAINS 1 SUSHI (SCR) DOMAIN.
CC -1- SIMILARITY: BELONGS TO THE AGGREGAN/VERSICAN PROTEOGLYCAN FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>)

or send an email to license@isb-sib.ch.

CC EMBL; U16306; AAA65018.1; -
CC EMBL; X15998; CAA34128.1; -
CC EMBL; S52488; AAB24878.1; -
CC EMBL; U26555; AAA67565.1; -
CC EMBL; D32039; BAA06801.1; -
CC EMBL; J02814; AAA36437.1; -
CC EMBL; AF084545; AAD48545.1; -
CC PIR; S06014; S06014.
CC PIR; A29348; A29348.
CC PIR; A30358; A30358.
CC HSSP; P01132; LEGF.
CC Genew; HGNC:2464; CSPG2.
CC MIM; 118661; -
CC InterPro; IPR000152; Asx_hydroxyl.
CC InterPro; IPR000561; EGF-like.
CC InterPro; IPR000742; EGF_2.
CC InterPro; IPR001881; EGF_Ca.
CC InterPro; IPR001438; EGF_II.
CC InterPro; IPR003599; Ig.
CC InterPro; IPR003006; Ig_MHC.
CC InterPro; IPR001304; Lectin_C.
CC InterPro; IPR000538; Link.
CC InterPro; IPR000436; Sushi_SCR_CCP.
CC Pfam; PF00008; EGF; 2.
CC Pfam; PF00047; Ig; 1.
CC Pfam; PF00059; lectin_c; 1.
CC Pfam; PF00084; sushi; 1.
CC Pfam; PF00193; xlink; 2.
CC PRINTS; PR00010; EGFBL00D.
CC ProDom; PD000918; Link; 2.
CC SMART; SM00032; CCP; 1.
CC SMART; SM00034; CLECT; 1.
CC SMART; SM00179; EGF_CA; 1.
CC SMART; SM00001; EGF_like; 1.
CC SMART; SM00409; IG; 1.
CC SMART; SM00445; LINK; 2.
CC PROSITE; PS00010; ASX_HYDROXYL; 1.
CC PROSITE; PS00022; EGF_1; 2.
CC PROSITE; PS01186; EGF_2; 1.
CC PROSITE; PS01187; EGF_CA; 1.
CC PROSITE; PS01241; LINK; 2.
CC PROSITE; PS00615; C_TYPE_LLECTIN_1; 1.
CC PROSITE; PS50041; C_TYPE_LLECTIN_2; 1.
CC Glycoprotein; Proteoglycan; Lectin; Extracellular matrix; Sushi;
CC Signal; Repeat; EGF-like domain; Calcium; Immunoglobulin domain;
CC Hyaluronic acid; Alternative splicing.
CC SIGNAL 1 20 POTENTIAL.
CC CHAIN 21 3396 VERSICAN CORE PROTEIN.
CC DOMAIN 37 137 IG-LIKE V-TYPE DOMAIN.
CC DOMAIN 167 244 LINK 1.
CC DOMAIN 265 346 LINK 2.
CC DOMAIN 348 1335 GAG-ALPHA (GLUCOSAMINOGLYCAN ATTACHMENT
CC DOMAIN).
CC DOMAIN 1336 3089 GAG-BETA.
CC DOMAIN 3089 3125 EGF-LIKE 1.
CC DOMAIN 3127 3163 EGF-LIKE 2, CALCIUM-BINDING (POTENTIAL).
CC DOMAIN 3176 3290 C-TYPE LECTIN.
CC DOMAIN 3295 3353 SUSHI.
CC DISULFID 44 130 BY SIMILARITY.
CC DISULFID 172 243 BY SIMILARITY.
CC DISULFID 196 217 BY SIMILARITY.
CC DISULFID 270 345 BY SIMILARITY.
CC DISULFID 294 315 BY SIMILARITY.
CC DISULFID 3093 3104 BY SIMILARITY.
CC DISULFID 3098 3113 BY SIMILARITY.
CC DISULFID 3115 3124 BY SIMILARITY.
CC DISULFID 3131 3142 BY SIMILARITY.
CC DISULFID 3136 3151 BY SIMILARITY.
CC DISULFID 3153 3162 BY SIMILARITY.
CC DISULFID 3169 3180 BY SIMILARITY.
CC DISULFID 3197 3289 BY SIMILARITY.

FT	DISULFID	3265	3281	BY SIMILARITY.	
FT	DISULFID	3296	3339	BY SIMILARITY.	
FT	DISULFID	3325	3352	BY SIMILARITY.	
FT	CARBOHYD	57	57	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	330	330	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	615	615	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	782	782	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	809	809	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	1332	1332	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	1398	1398	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	1442	1442	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	1468	1468	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	1663	1663	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	1898	1898	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	2179	2179	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	2272	2272	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	2280	2280	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	2360	2360	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	2385	2385	N-LINKED (GLCNAC. . .)	(POTENTIAL).

Query Match

Best Local Similarity

Matches 47; Conservative 23; Mismatches 55; Indels 40; Gaps 8;

15.6%; Score 174; DB 1; Length 3396;

28.5%; Pred. No. 3.9e-07;

QY	17	QRPCYKVIYFHTSRLNFEAEKACRRDGGQLVSIESEDEQKLEKFIENLLPSDGF-	75
Db	3177	QGQCYK--YF---AHRRTWDAARECRLOGAHLTSILSHEEQMFVNRV-----GHDIQ	3224
QY	76	WIGLRRREKQSNSTACQDLAYWTDGSIQFRNYYVDEP---SCGSEVCVVMYHQPSAP	131
Db	3225	WIGL-----NDKMFEDHDFRWDGSLQYENWRPNQDPSFSGEDCVVWIWHENG--	3274
QY	132	AGIGGPFYMFQWDDRCNMKNFNICKYS---DEKPAVPSREAEGE	172
Db	3275	-----QWNVPCNYHLTYTCKKGTACGQPPVVENAKTFGK	3310

RESULT 10

PGCV_CHICK

STANDARD; PRT; 3562 AA.

AC	Q90953; Q90945;
DT	01-NOV-1997 (Rel. 35, Created)
DT	01-NOV-1997 (Rel. 35, Last sequence update)
DT	15-JUN-2002 (Rel. 41, Last annotation update)
DE	Versican core protein precursor (Large fibroblast proteoglycan)
DE	(Chondroitin sulfate proteoglycan core protein 2) (PG-M).
GN	CPSG2.
OS	Gallus gallus (Chicken).
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC	Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC	Gallus.
OX	NCBI_TaxID=9031;
RN	[1]
RP	SEQUENCE FROM N.A.
RC	STRAIN-White leghorn; TISSUE-Limb bud;
RX	MEDLINE-93300846; PubMed-8314802;
RA	Shinomura T., Nishida Y., Ito K., Kimata K.;
RT	"CDNA cloning of PG-M, a large chondroitin sulfate proteoglycan
RT	expressed during chondrogenesis in chick limb buds. Alternative
RT	spliced multiforms of PG-M and their relationships to versican.";
RL	J. Biol. Chem. 268:14461-14469(1993).
CC	-1- FUNCTION: May play a role in intercellular signaling and in
CC	connecting cells with the extracellular matrix. May take part in
CC	the regulation of cell motility, growth and differentiation. Binds
CC	hyaluronic acid.
CC	-1- SUBCELLULAR LOCATION: Secreted; extracellular matrix.
CC	-1- ALTERNATIVE PRODUCTS: At least 2 isoforms; V0 (shown here) and
CC	V1; are produced by alternative splicing.
CC	-1- TISSUE SPECIFICITY: Prechondrogenic condensation area of
CC	developing limb buds.
CC	-1- DEVELOPMENTAL STAGE: Disappears after the cartilage development
CC	(By similarity).
CC	-1- SIMILARITY: CONTAINS 1 IMMUNOGLOBULIN-LIKE V-TYPE DOMAIN

CC	-1- SIMILARITY: CONTAINS 2 LINK DOMAINS.				
CC	-1- SIMILARITY: CONTAINS 2 EGF-LIKE DOMAINS.				
CC	-1- SIMILARITY: CONTAINS 1 C-TYPE LECTIN FAMILY DOMAIN.				
CC	-1- SIMILARITY: CONTAINS 1 SUSHI (SCR) DOMAIN.				
CC	-----				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).				
CC	-----				
CC	EMBL; X60226; CAA42787.1; -.				
DR	EMBL; D13542; BAA02742.1; -.				
DR	HSSP; P00740; IEDM.				
DR	InterPro; IPR000152; Asx_hydroxyl.				
DR	InterPro; IPR000561; EGF-like.				
DR	InterPro; IPR000742; EGF_2.				
DR	InterPro; IPR001881; EGF_Ca.				
DR	InterPro; IPR003599; Ig.				
DR	InterPro; IPR003006; Ig_MHC.				
DR	InterPro; IPR001304; Lectin_C.				
DR	InterPro; IPR000538; Link.				
DR	InterPro; IPR000436; Sushi_SCR_CCP.				
DR	Pfam; PF00008; EGF; 2.				
DR	Pfam; PF00047; Ig; 1.				
DR	Pfam; PF00059; lectin_c; 1.				
DR	Pfam; PF00084; sushi; 1.				
DR	Pfam; PF00193; Xlink; 2.				
DR	ProDom; PD000918; Link; 2.				
DR	SMART; SM00032; CCP; 1.				
DR	SMART; SM00034; CLECT; 1.				
DR	SMART; SM00179; EGF_CA; 1.				
DR	SMART; SM00001; EGF_like; 1.				
DR	SMART; SM00409; IG; 1.				
DR	SMART; SM00445; LINK; 2.				
DR	PROSITE; PS00010; ASX_HYDROXYL; 1.				
DR	PROSITE; PS00022; EGF_1; 2.				
DR	PROSITE; PS01186; EGF_2; 1.				
DR	PROSITE; PS01187; EGF_CA; 1.				
DR	PROSITE; PS01241; LINK; 2.				
DR	PROSITE; PS00615; C_TYPE_LECTIN_1; 1.				
DR	PROSITE; PS50041; C_TYPE_LECTIN_2; 1.				
KW	Glycoprotein; Proteoglycan; Lectin; Extracellular matrix; Sushi;				
KW	Signal; Repeat; EGF-like domain; Calcium; Immunoglobulin domain;				
KW	Hyaluronic acid; Alternative splicing.				
FT	SIGNAL	1	26	POTENTIAL.	
FT	CHAIN	27	3562	VERSICAN CORE PROTEIN.	
FT	DOMAIN	37	136	IG-LIKE V-TYPE DOMAIN.	
FT	DOMAIN	166	243	LINK 1.	
FT	DOMAIN	264	345	LINK 2.	
FT	DOMAIN	3254	3290	EGF-LIKE 1.	
FT	DOMAIN	3292	3328	EGF-LIKE 2, CALCIUM-BINDING (POTENTIAL).	
FT	DOMAIN	3341	3455	C-TYPE LECTIN.	
FT	DOMAIN	3460	3518	SUSHI.	
FT	DISULFID	44	129	BY SIMILARITY.	
FT	DISULFID	171	242	BY SIMILARITY.	
FT	DISULFID	195	216	BY SIMILARITY.	
FT	DISULFID	269	344	BY SIMILARITY.	
FT	DISULFID	293	314	BY SIMILARITY.	
FT	DISULFID	3258	3269	BY SIMILARITY.	
FT	DISULFID	3263	3278	BY SIMILARITY.	
FT	DISULFID	3280	3289	BY SIMILARITY.	
FT	DISULFID	3296	3307	BY SIMILARITY.	
FT	DISULFID	3301	3316	BY SIMILARITY.	
FT	DISULFID	3318	3327	BY SIMILARITY.	
FT	DISULFID	3334	3345	BY SIMILARITY.	
FT	DISULFID	3362	3454	BY SIMILARITY.	
FT	DISULFID	3430	3446	BY SIMILARITY.	
FT	DISULFID	3461	3504	BY SIMILARITY.	
FT	DISULFID	3490	3517	BY SIMILARITY.	
FT	CARBOHYD	163	163	N-LINKED (GLCNAC. . .) (POTENTIAL).	

FT	CARBOHYD	235	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	329	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	529	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	709	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	948	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	1409	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	1479	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	1523	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	1530	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	1625	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	1751	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	1988	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	2088	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	2089	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	2507	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	2642	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	2679	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	2748	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	2762	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	3069	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	3194	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	3232	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	3545	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	VARSPLIC	485	MISSING (IN ISOFORM V1).	
SQ	SEQUENCE	3562 AA; 388078 MW; 9BC566E88C1602D2	CRC64;	
Query Match 15.3%; Score 171; DB 1; Length 3562;				
Best Local Similarity 28.5%; Pred. No. 7.6e-07;				
Matches 47; Conservative 23; Mismatches 55; Indels 40; Gaps 8;				
QY	17	QRPCYKVIYFHDTSRLNFEEAKEACRRDGGQLVSISEDEQKLIKFIENLLPSDGF-	75	
Db	3342	QGQCYK--YF---AHRRTWDTAERECLQGAHLTSILSHEEQVFNRI-----GHDYQ	3389	
QY	76	WIGLRREEKQSNSTACQDLYAWTDGSIQFRNMYVDEP---SCGSEVCVVMYHQPSAP	131	
Db	3390	WIGL-----NDKMFERDFRTDGSPLQYENWRPNQDFFSAGEDCVVIWHENG--	3439	
QY	132	AGIGPYMFQWDDRCNMKNFICKYS----DEKPAVPSREAEGE	172	
Db	3440	-----QWNDVPCNVHLTYTCKKGTVACGPPVVENAKTFGK	3475	
RESULT 11				
CD93_RAT	STANDARD;	PRT;	643 AA.	
ID	CD93_RAT	Q9ET61; Q9J126;		
AC	15-JUN-2002 (Rel. 41, Created)			
DT	15-JUN-2002 (Rel. 41, Last sequence update)			
DT	15-JUN-2002 (Rel. 41, Last annotation update)			
DE	Complement component C1q receptor precursor (Complement component 1, q			
DE	subcomponent, receptor 1) (C1qRp) (C1qRp) (C1q/MBL/SPA receptor)			
DE	(CD93 antigen) (Cell surface antigen AA4).			
GN	C1QR1 OR CD93 OR C1QRP.			
OS	Rattus norvegicus (Rat).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Euthera; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.			
OX	NCBI_TaxID=10116;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=PVG; TISSUE=Natural killer cells;			
RX	MEDLINE=20545218; Pubmed=11093152;			
RA	Lovik G., Vaage J.T., Dissen E., Szpirer C., Ryan J.C., Rolstad B.;			
RT	"Characterization and molecular cloning of rat C1qRp, a receptor on NK			
RT	cells.";			
RL	Eur. J. Immunol. 30:3355-3362(2000).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=Wistar; TISSUE=Lung;			
RX	MEDLINE=20507883; Pubmed=10934210;			
RA	Dean Y.D., McGreal E.P., Akatsu H., Gasque P.;			
RT	"Molecular and cellular properties of the rat AA4 antigen, a C-type			
RT	lectin-like receptor with structural homology to thrombomodulin.";			

RL	J. Biol. Chem. 275:34382-34392(2000).			
CC	-1- FUNCTION: Receptor (or element of a larger receptor complex) for			
CC	C1q, mannose-binding lectin (MBL2) and pulmonary surfactant			
CC	protein A (SPA). May mediate the enhancement of phagocytosis in			
CC	monocytes and macrophages upon interaction with soluble defense			
CC	collagens. May play a role in intercellular adhesion.			
CC	-1- SUBCELLULAR LOCATION: Type I membrane protein.			
CC	-1- TISSUE SPECIFICITY: Widely expressed. Highly expressed in lung and			
CC	heart. Expressed at lower level in brain, thymus, liver, spleen,			
CC	intestine, kidney, adrenal gland, muscle and testis. Expressed on			
CC	endothelial cells, platelets, undifferentiated monocytes and			
CC	circulating natural killer cells.			
CC	-1- PTM: N- and O-glycosylated (By similarity).			
CC	-1- SIMILARITY: CONTAINS 1 C-TYPE LECTIN FAMILY DOMAIN.			
CC	-1- SIMILARITY: CONTAINS 5 EGF-LIKE DOMAINS.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb-sib.ch/			
CC	or send an email to license@isb-sib.ch).			
CC	-----			
DR	EMBL; AF136537; AAG01572.1; -.			
DR	HSSP; P35555; 1EMN.			
DR	InterPro; IPR000152; Asx_hydroxyl.			
DR	InterPro; IPR000561; EGF-like.			
DR	InterPro; IPR001881; EGF_Ca.			
DR	InterPro; IPR001304; Lectin_C.			
DR	Pfam; PF00008; EGF; 5.			
DR	Pfam; PF00059; lectin_c; 1.			
DR	SMART; SM00034; CLECT; 1.			
DR	SMART; SM00181; EGF; 5.			
DR	SMART; SM00179; EGF_CA; 5.			
DR	SMART; SM00001; EGF_Like; 2.			
DR	PROSITE; PS00010; ASX_HYDROXYL; 3.			
DR	PROSITE; PS00615; C_TYPE_LECTIN_1; FALSE_NEG.			
DR	PROSITE; PS50041; C_TYPE_LECTIN_2; 1.			
DR	PROSITE; PS01186; EGF_2; 3.			
DR	PROSITE; PS01187; EGF_CA; 3.			
KW	Receptor; EGF-like domain; Signal; Transmembrane; Glycoprotein;			
KW	Repeat; Lectin.			
FT	SIGNAL	1	23	POTENTIAL.
FT	CHAIN	24	643	COMPLEMENT COMPONENT C1Q RECEPTOR.
FT	DOMAIN	24	571	EXTRACELLULAR (POTENTIAL).
FT	TRANSMEM	572	592	POTENTIAL.
FT	DOMAIN	593	643	CYTOPLASMIC (POTENTIAL).
FT	DOMAIN	31	173	C-TYPE LECTIN.
FT	DOMAIN	257	298	EGF-LIKE 1.
FT	DOMAIN	299	341	EGF-LIKE 2.
FT	DOMAIN	342	381	EGF-LIKE 3.
FT	DOMAIN	382	423	EGF-LIKE 4.
FT	DOMAIN	424	462	EGF-LIKE 5.
FT	DISULFID	261	272	BY SIMILARITY.
FT	DISULFID	268	282	BY SIMILARITY.
FT	DISULFID	284	297	BY SIMILARITY.
FT	DISULFID	303	314	BY SIMILARITY.
FT	DISULFID	308	325	BY SIMILARITY.
FT	DISULFID	327	340	BY SIMILARITY.
FT	DISULFID	346	355	BY SIMILARITY.
FT	DISULFID	351	364	BY SIMILARITY.
FT	DISULFID	366	380	BY SIMILARITY.
FT	DISULFID	386	397	BY SIMILARITY.
FT	DISULFID	393	406	BY SIMILARITY.
FT	DISULFID	408	422	BY SIMILARITY.
FT	DISULFID	428	437	BY SIMILARITY.
FT	DISULFID	433	446	BY SIMILARITY.
FT	DISULFID	448	461	BY SIMILARITY.
FT	CARBOHYD	322	322	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	498	498	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CONFLICT	417	417	E -> K (IN REF. 2).

CC IS TO RESIST COMPRESSION IN CARTILAGE. IT BINDS AVIDLY TO
CC HYALURONIC ACID VIA AN AMINO-TERMINAL GLOBULAR REGION. MAY PLAY A
CC REGULATORY ROLE IN THE MATRIX ASSEMBLY OF THE CARTILAGE.
CC -1- SUBCELLULAR LOCATION: Secreted; extracellular matrix (By
CC similarity).

CC -1- DOMAIN: TWO GLOBULAR DOMAINS, G1 AND G2, COMPRISE THE AMINO
CC TERMINUS OF THE PROTEOGLYCAN, WHILE ANOTHER GLOBULAR REGION, G3,
CC MAKES UP THE COOH TERMINUS. G1 CONTAINS LINK DOMAINS AND THUS
CC CONSISTS OF THREE DISULFIDE-BONDED LOOP STRUCTURES DESIGNATED AS
CC THE A, B, B' MOTIFS. G2 IS SIMILAR TO G1. THE KERATAN SULFATE (KS)
CC AND THE CHONDROITIN SULFATE (CS) ATTACHMENT DOMAINS LIE BETWEEN G2
CC AND G3.

CC -1- PTM: CONTAINS MOSTLY CHONDROITIN SULFATE, BUT ALSO KERATAN SULFATE
CC CHAINS, N-LINKED AND O-LINKED OLIGOSACCHARIDES (BY SIMILARITY).

CC -1- SIMILARITY: CONTAINS 1 IMMUNOGLOBULIN-LIKE V-TYPE DOMAIN.

CC -1- SIMILARITY: CONTAINS 4 LINK DOMAINS.

CC -1- SIMILARITY: CONTAINS 1 C-TYPE LECTIN FAMILY DOMAIN.

CC -1- SIMILARITY: CONTAINS 1 SUSHI (SCR) DOMAIN.

CC -1- SIMILARITY: CONTAINS 1 EGF-LIKE DOMAIN.

CC -1- SIMILARITY: BELONGS TO THE AGGREGAN/VERSICAN PROTEOGLYCAN FAMILY.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC -----

DR EMBL; U65989; AAB06238.2; -;
DR EMBL; S74662; AAC60527.1; -;
DR EMBL; L07054; -; NOT_ANNOTATED_CDS.
DR HSSP; P08709; 1BF9.
DR InterPro; IPR000152; Asx_hydroxyl.
DR InterPro; IPR000561; EGF-like.
DR InterPro; IPR000742; EGF 2.
DR InterPro; IPR001881; EGF_Ca.
DR InterPro; IPR003599; Ig.
DR InterPro; IPR003006; Ig_MHC.
DR InterPro; IPR001304; Lectin_C.
DR InterPro; IPR000538; Link.
DR InterPro; IPR003324; SGXXSG.
DR InterPro; IPR000436; Sushi_SCR_CCP.
DR Pfam; PF00008; EGF; 1.
DR Pfam; PF00047; Ig; 1.
DR Pfam; PF00059; lectin_c; 1.
DR Pfam; PF00084; sushi; 1.
DR Pfam; PF00193; Xlink; 4.
DR Pfam; PF02339; SGXXSG; 70.
DR PRINTS; PR01265; LINKMODULE.
DR ProDom; PD000918; Link; 4.
DR SMART; SM00032; CCP; 1.
DR SMART; SM00034; CLECT; 1.
DR SMART; SM00179; EGF_Ca; 1.
DR SMART; SM00409; Ig; 1.
DR SMART; SM00445; LINK; 4.
DR PROSITE; PS00010; ASX_HYDROXYL; 1.
DR PROSITE; PS00022; EGF_1; 1.
DR PROSITE; PS01187; EGF_Ca; 1.
DR PROSITE; PS00290; IG_MHC; 1.
DR PROSITE; PS01241; LINK; 4.
DR PROSITE; PS00615; C_TYPE_LLECTIN_1; 1.
DR PROSITE; PS50041; C_TYPE_LLECTIN_2; 1.
KW Glycoprotein; Cartilage; Proteoglycan; Lectin; Signal; Sushi;
KW EGF-like domain; Repeat; Immunoglobulin domain.
FT SIGNAL 1 16 POTENTIAL.
FT CHAIN 17 2333 AGGREGAN CORE PROTEIN.
FT DOMAIN 44 140 IG-LIKE V-TYPE DOMAIN.
FT DOMAIN 170 247 LINK 1.
FT DOMAIN 268 349 LINK 2.
FT DOMAIN 513 590 LINK 3.
FT DOMAIN 611 692 LINK 4.
FT DOMAIN 2081 2117 EGF-LIKE, CALCIUM-BINDING (POTENTIAL).

FT DOMAIN 2130 2245 C-TYPE LECTIN.
FT DOMAIN 2249 2307 SUSHI.
FT DOMAIN 48 140 G1-A.
FT DOMAIN 152 247 G1-B.
FT DOMAIN 253 349 G1-B'.
FT DOMAIN 495 589 G2-B.
FT DOMAIN 596 691 G2-B'.
FT DOMAIN 694 816 KS.
FT DOMAIN 819 1394 CS-1.
FT DOMAIN 1395 2079 CS-2.
FT DOMAIN 2080 2333 G3.
FT DISULFID 51 133 BY SIMILARITY.
FT DISULFID 175 246 BY SIMILARITY.
FT DISULFID 199 220 BY SIMILARITY.
FT DISULFID 273 348 BY SIMILARITY.
FT DISULFID 297 318 BY SIMILARITY.
FT DISULFID 518 589 BY SIMILARITY.
FT DISULFID 542 563 BY SIMILARITY.
FT DISULFID 616 691 BY SIMILARITY.
FT DISULFID 640 661 BY SIMILARITY.
FT DISULFID 2085 2096 BY SIMILARITY.
FT DISULFID 2090 2105 BY SIMILARITY.
FT DISULFID 2107 2116 BY SIMILARITY.
FT DISULFID 2123 2134 BY SIMILARITY.
FT DISULFID 2151 2243 BY SIMILARITY.
FT DISULFID 2219 2235 BY SIMILARITY.
FT DISULFID 2250 2293 BY SIMILARITY.
FT DISULFID 2279 2306 BY SIMILARITY.
FT CARBOHYD 126 126 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 239 239 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 333 333 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 387 387 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 444 444 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 620 620 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 676 676 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 747 747 N-LINKED (GLCNAC...) (POTENTIAL).
SQ SEQUENCE 2333 AA; 240573 MW; 8B9ED78F3508B596 CRC64;
Query Match 13.9%; Score 155.5; DB 1; Length 2333;
Best Local Similarity 28.5%; Pred. No. 1.1e-05;
Matches 51; Conservative 24; Mismatches 61; Indels 43; Gaps 11;
QY 9 QPVCRRG---TQPCYKVIYFHDTSRRLNFEAEKACRRDGGQLVSIESEDEQKLEKFI 65
| : | | | | | | | | | : : : | | | | | : | | : | |
Db 2120 QELCEEGWTKFQGHCHYR--YFPD---RESWVDAESRCRAQQSHLSSIVTPEEQ----EFV 2170
QY 66 ENLLPSDGDGFWIGLRRREKQSNSTACQDLYAWTDGSIQFRNMYVDEPS---CGSEVC 121
| : | : | | | | | : : : | | | | | : | | : | |
Db 2171 NN---NAQDYQWIGL-----NDRTEIGDFRWSGDHSLQFENWRPNQPDNFFVSGEDC 2219
QY 122 VVM-YHQPSAPAGIGGPFYMFQWNDRCNMKNNFICKYS----DEKPAVPSREAEGETE 175
| | | : | : | : | | | | | : | | : | | : | : | :
Db 2220 VVMWHEKG-----EWNDVPCNYLPFTCKKGTVACGDPVVEHARTFGQKDD 2267
RESULT 15
LEM2_MOUSE
ID LEM2_MOUSE STANDARD; PRT; 612 AA.
AC Q00690;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE E-selectin precursor (Endothelial leukocyte adhesion molecule 1)
DE (ELAM-1) (Leukocyte-endothelial cell adhesion molecule 2) (LECAM2)
DE (CD62E).
GN SELE OR ELAM-1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92283265; PubMed=1375914;

RA Becker-Andre M., van Huijsduijn R.H., Losberger C., Whelan J.,
RA Delamarter J.F.;
RT "Murine endothelial leukocyte-adhesion molecule 1 is a close
RT structural and functional homologue of the human protein.";
RL Eur. J. Biochem. 206:401-411(1992).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE-92340571; PubMed-1378846;
RA Weller A., Isenmann S., Vestweber D.;
RT "Cloning of the mouse endothelial selectins. Expression of both E-
RT and P-selectin is inducible by tumor necrosis factor alpha.";
RL J. Biol. Chem. 267:15176-15183(1992).
CC -1- FUNCTION: EXPRESSED ON CYTOKINE INDUCED ENDOTHELIAL CELLS AND
CC MEDIATES THEIR BINDING TO LEUKOCYTES. THE LIGAND RECOGNIZED BY
CC ELAM-1 IS SIALYL-LEWIS X (ALPHA(1->3)FUCOSYLATED DERIVATIVES OF
CC POLYLACTOSAMINE THAT ARE FOUND AT THE NONREDUCING TERMINI OF
CC GLYCOLIPIDS).
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- SIMILARITY: BELONGS TO THE SELECTIN/LECAM FAMILY.
CC -1- SIMILARITY: CONTAINS 1 C-TYPE LECTIN FAMILY DOMAIN.
CC -1- SIMILARITY: CONTAINS 1 EGF-LIKE DOMAIN.
CC -1- SIMILARITY: CONTAINS 6 SUSHI (SCR) DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M80778; AAA37547.1; -
DR EMBL; M87862; AAA37577.1; ALT_INIT.
DR HSSP; P16581; 1KJA.
DR MGD; MGI:98278; Sele.
DR InterPro; IPR000561; EGF-like.
DR InterPro; IPR001304; Lectin_C.
DR InterPro; IPR002396; Selectin.
DR InterPro; IPR000436; Sushi_SCR_CCP.
DR Pfam; PF00008; EGF; 1.
DR Pfam; PF00059; lectin_c; 1.
DR Pfam; PF00084; sushi; 6.
DR PRINTS; PR00343; SELECTIN.
DR SMART; SM00032; CCP; 6.
DR SMART; SM00034; CLECT; 1.
DR SMART; SM00181; EGF; 1.
DR PROSITE; PS00022; EGF_1; 1.
DR PROSITE; PS01186; EGF_2; 1.
DR PROSITE; PS00615; C_TYPE_LECTIN_1; 1.
DR PROSITE; PS50041; C_TYPE_LECTIN_2; 1.
KW Cell adhesion; Transmembrane; Glycoprotein; EGF-like domain; Lectin;
KW Selectin; Signal; Sushi; Repeat.
FT SIGNAL 1 21 BY SIMILARITY.
FT CHAIN 22 612 E-SELECTIN.
FT DOMAIN 22 557 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 558 579 POTENTIAL.
FT DOMAIN 580 612 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 38 138 C-TYPE LECTIN (SHORT FORM).
FT DOMAIN 139 175 EGF-LIKE.
FT DOMAIN 179 239 SUSHI 1.
FT DOMAIN 242 301 SUSHI 2.
FT DOMAIN 304 364 SUSHI 3.
FT DOMAIN 367 427 SUSHI 4.
FT DOMAIN 430 490 SUSHI 5.
FT DOMAIN 493 549 SUSHI 6.
FT DISULFID 40 138 BY SIMILARITY.
FT DISULFID 111 130 BY SIMILARITY.
FT DISULFID 143 154 BY SIMILARITY.
FT DISULFID 148 163 BY SIMILARITY.
FT DISULFID 165 174 BY SIMILARITY.
FT DISULFID 180 225 BY SIMILARITY.
FT DISULFID 210 238 BY SIMILARITY.
FT DISULFID 243 287 BY SIMILARITY.

FT	DISULFID	273	300	BY SIMILARITY.
FT	DISULFID	305	350	BY SIMILARITY.
FT	DISULFID	336	363	BY SIMILARITY.
FT	DISULFID	368	413	BY SIMILARITY.
FT	DISULFID	399	426	BY SIMILARITY.
FT	DISULFID	431	476	BY SIMILARITY.
FT	DISULFID	462	489	BY SIMILARITY.
FT	DISULFID	494	535	BY SIMILARITY.
FT	DISULFID	521	548	BY SIMILARITY.
FT	CARBOHYD	25	25	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	145	145	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	192	192	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	203	203	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	266	266	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	333	333	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	391	391	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	504	504	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	528	528	N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ	SEQUENCE	612 AA; 66749 MW; 86F05713F0EC2C3D CRC64;		

Query Match 13.9%; Score 154.5; DB 1; Length 612;
Best Local Similarity 27.9%; Pred. No. 2.8e-06;
Matches 38; Conservative 28; Mismatches 45; Indels 25; Gaps 5;

QY	25	YFHDTSRRLNFEAEACRRDGGQLVSISEDEQKLIKFIENLLPSDGFWIGLRRREE	84
Db	23	YYNASSELMTYDEASAYCQRDYTHLVAIQNKEE---INVLNSNLKHSPSYWIGIRK---	76
QY	85	KQSNSTACQDLAYAWTDGS---ISQFRNWWYVDEPS--CGSEVCVVMYHQPSAPAGIGGPYM	139
Db	77	-----VNNVWIWVGTKPLTEEAQNWAPGEPNNKORNEDCIVEIYIORTKDSGM-----	124
QY	140	FQWDDRCNMKNFIC	155
Db	125	--WDERCNKKKLAIC	138

Search completed: May 20, 2003, 17:05:30
Job time : 19 secs

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 20, 2003, 17:02:46 ; Search time 31 Seconds
(without alignments)
1369.216 Million cell updates/sec

Title: US-09-887-855-5
Perfect score: 1115
Sequence: 1 ATGRLLSGQPVCRGGTQRPC.....EEDAKKTFKESREALNLAY 206

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

- Database : SPTREMBL_21.*
- 1: sp_archaea.*
 - 2: sp_bacteria.*
 - 3: sp_fungi.*
 - 4: sp_human.*
 - 5: sp_invertebrate.*
 - 6: sp_mammal.*
 - 7: sp_mhc.*
 - 8: sp_organelle.*
 - 9: sp_phage.*
 - 10: sp_plant.*
 - 11: sp_rodent.*
 - 12: sp_virus.*
 - 13: sp_vertebrate.*
 - 14: sp_unclassified.*
 - 15: sp_rvirus.*
 - 16: sp_bacteriap.*
 - 17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1115	100.0	374	4 Q96NF3	Q96nf3 homo sapien
2	1115	100.0	374	4 Q8TAY8	Q8tay8 homo sapien
3	1111	99.6	374	4 Q96NC5	Q96nc5 homo sapien
4	945.5	84.8	374	11 Q92209	Q92209 cricetulus
5	574	51.5	273	11 Q8VI31	Q8vi31 mus musculu
6	185	16.6	1290	13 Q9W6E1	Q9w6el gallus gall
7	178.5	16.0	1456	11 Q61830	Q61830 mus musculu
8	177.5	15.9	1348	5 Q25199	Q25199 hydra atten
9	176.5	15.8	1479	4 Q9Y5P9	Q9y5p9 homo sapien
10	176.5	15.8	1479	4 Q9UBG0	Q9ubg0 homo sapien
11	176	15.8	315	4 Q9UPK6	Q9upk6 homo sapien
12	176	15.8	1321	4 O14594	O14594 homo sapien
13	170.5	15.3	1479	11 Q64449	Q64449 mus musculu
14	161.5	14.5	134	5 Q9XYX3	Q9xyx3 hydra magni
15	159.5	14.3	742	11 Q8VIF6	Q8vif6 mus musculu
16	158.5	14.2	719	6 O62623	O62623 bos taurus

17	158	14.2	295	11 Q912W4	Q91zw4 mus musculu
18	158	14.2	311	11 Q9D8V4	Q9d8v4 mus musculu
19	158	14.2	325	11 Q912X0	Q91zx0 mus musculu
20	157	14.1	158	13 Q90WI7	Q90wi7 bungarus fa
21	156	14.0	339	6 Q95244	Q95244 sus scrofa
22	155	13.9	1152	13 Q90WM2	Q90wm2 xenopus lae
23	152.5	13.7	485	6 Q95LG3	Q95lg3 odocoileus
24	152	13.6	158	13 Q90WI6	Q90wi6 bungarus mu
25	150	13.5	197	6 Q28008	Q28008 bos taurus
26	149.5	13.4	459	5 Q22136	Q22136 caenorhabdi
27	149	13.4	158	13 Q90WI8	Q90wi8 bungarus fa
28	148.5	13.3	195	5 Q27340	Q27340 megabalanus
29	148.5	13.3	330	4 Q9NT67	Q9nt67 homo sapien
30	148.5	13.3	911	4 Q9HBK4	Q9hbk4 homo sapien
31	148.5	13.3	911	4 Q96GW7	Q96gw7 homo sapien
32	148.5	13.3	911	4 Q96FP7	Q96fp7 homo sapien
33	148.5	13.3	911	4 Q8TBB9	Q8tbb9 homo sapien
34	148	13.3	253	5 Q20665	Q20665 caenorhabdi
35	148	13.3	381	6 Q8SQB2	Q8sqb2 macaca mula
36	147	13.2	196	11 Q9EPW4	Q9epw4 mus musculu
37	147	13.2	381	6 Q95LC6	Q95lc6 macaca neme
38	147	13.2	404	6 Q95J96	Q95j96 macaca mula
39	146.5	13.1	152	13 Q9DG39	Q9dg39 agkistrodon
40	146.5	13.1	162	5 Q25459	Q25459 megabalanus
41	146.5	13.1	1031	5 Q8WSX2	Q8wsx2 dugesia tig
42	146	13.1	2109	13 P79787	P79787 gallus gall
43	145.5	13.0	134	5 Q9UB05	Q9ub05 hydra atten
44	145.5	13.0	381	6 Q95LA8	Q95la8 macaca mula
45	145.5	13.0	404	6 Q95L98	Q95l98 pan troglod

ALIGNMENTS

RESULT 1
Q96NF3 PRELIMINARY; PRT; 374 AA.
AC Q96NF3;
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE CDNA FLJ30977 fis, clone HHDPC2000095, highly similar to Cricetulus griseus layilin mRNA.
DE griseus layilin mRNA.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Ishibashi T., Kanehori K., Yosida M., Watanabe S., Ishida S., Ono Y., RA Hotuta T., Hiraoka S., Murakawa K., Takiguchi S., Kusano J., Chiba Y., RA Watanabe M., Fujimori K., Tanai H., Ishida M., Yamashita H., Chiba Y., RA Sugiyama T., Irie R., Otsuki T., Sato H., Wakamatsu A., Ishii S., RA Yamamoto J., Isono Y., Kawai-Hio Y., Saito K., Nishikawa T., RA Kimura K., Matsuo K., Nakamura Y., Sekine M., Kikuchi H., Kanda K., RA Wagatsuma M., Takahashi-Fujii A., Oshima A., Sugiyama A., Kawakami B., RA Suzuki Y., Sugano S., Nagahari K., Masuho Y., Nagai K., Isogai T.; RT "NEDO human cDNA sequencing project."; RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases. DR EMBL; AK055539; BAB70946.1; -. DR InterPro; IPR001304; LECTIN_C. DR Pfam; PF00059; lectin_c; 1. DR PROSITE; PS50041; C_TYPE_LLECTIN_2; 1. SQ SEQUENCE 374 AA; 42280 MW; 8AE64E6BC9E56DCD CRC64;

Query Match 100.0%; Score 1115; DB 4; Length 374;
Best Local Similarity 100.0%; Pred. No. 3.8e-98;
Matches 206; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGRLLSGQPVCRGGTQRPCYKVIYFHDTSRRLNFEAKEACRRDGGQLVSISEDEQKL 60
Db 22 ATGRLLSGQPVCRGGTQRPCYKVIYFHDTSRRLNFEAKEACRRDGGQLVSISEDEQKL 81

QY 61 IEKFIENLLPSDGDWIGLRRREKQSNSTACQDLYAWTDGSIQFRNWWYVDEPSCGSEV 120
Db 82 IEKFIENLLPSDGDWIGLRRREKQSNSTACQDLYAWTDGSIQFRNWWYVDEPSCGSEV 141
QY 121 CVVMYHQPSAPAGIGGYPYMFQWDDRCNMKNFICKYSDEKPAVPSREAEGEETELTPV 180
Db 142 CVVMYHQPSAPAGIGGYPYMFQWDDRCNMKNFICKYSDEKPAVPSREAEGEETELTPV 201
QY 181 LPEETQEDAKKTFKESREAALNLAY 206
Db 202 LPEETQEDAKKTFKESREAALNLAY 227

RESULT 2
Q8TAY8
ID Q8TAY8 PRELIMINARY; PRT; 374 AA.
AC Q8TAY8;
DT 01-JUN-2002 (TREMBlrel. 21, Created)
DT 01-JUN-2002 (TREMBlrel. 21, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Similar to unnamed protein product.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=BRAIN;
RA Strausberg R.;
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC025407; AAH25407.1; --
SQ SEQUENCE 374 AA; 42312 MW; FC214E6BC9E578D9 CRC64;

Query Match 100.0%; Score 1115; DB 4; Length 374;
Best Local Similarity 100.0%; Pred. No. 3.8e-98;
Matches 206; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGRLLSGQPVCRCGGTQPCYKVIYFHDTSRRLNFEEAKEACRRDGGQLVSI ESEDEQKL 60
Db 22 ATGRLLSGQPVCRCGGTQPCYKVIYFHDTSRRLNFEEAKEACRRDGGQLVSI ESEDEQKL 81
QY 61 IEKFIENLLPSDGDWIGLRRREKQSNSTACQDLYAWTDGSIQFRNWWYVDEPSCGSEV 120
Db 82 IEKFIENLLPSDGDWIGLRRREKQSNSTACQDLYAWTDGSIQFRNWWYVDEPSCGSEV 141
QY 121 CVVMYHQPSAPAGIGGYPYMFQWDDRCNMKNFICKYSDEKPAVPSREAEGEETELTPV 180
Db 142 CVVMYHQPSAPAGIGGYPYMFQWDDRCNMKNFICKYSDEKPAVPSREAEGEETELTPV 201
QY 181 LPEETQEDAKKTFKESREAALNLAY 206
Db 202 LPEETQEDAKKTFKESREAALNLAY 227

RESULT 3
Q96NC5
ID Q96NC5 PRELIMINARY; PRT; 374 AA.
AC Q96NC5;
DT 01-DEC-2001 (TREMBlrel. 19, Created)
DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE CDNA FLJ31092 fis, clone IMR321000158, highly similar to Cricetulus griseus layilin.
DE griseus layilin.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Tashiro H., Yamazaki M., Watanabe K., Kumagai A., Itakura S., Fukuzumi Y., Fujimori Y., Komiyama M., Sugiyama T., Irie R., Otsuki T., Sato H., Wakamatsu A., Ishii S., Yamamoto J., Isono Y., Kawai-Hio Y., Saito K., Nishikawa T., Kimura K., Yamashita H.,

RA Matsuo K., Nakamura Y., Sekine M., Kikuchi H., Kanda K., Wagatsuma M.,
RA Murakawa K., Kanehori K., Takahashi-Fujii A., Oshima A., Sugiyama A.,
RA Kawakami B., Suzuki Y., Sugano S., Nagahari K., Masuho Y., Nagai K.,
RA Isogai T.;
RT "NEDO human cDNA sequencing project.";
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK055654; BAB70978.1; --
DR InterPro; IPR001304; Lectin_C.
DR Pfam; PF00059; lectin_c; 1.
DR PROSITE; PS50041; C_TYPE_LLECTIN_2; 1.
SQ SEQUENCE 374 AA; 42310 MW; CBF74E676E23BA16 CRC64;

Query Match 99.6%; Score 1111; DB 4; Length 374;
Best Local Similarity 99.5%; Pred. No. 9.2e-98;
Matches 205; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ATGRLLSGQPVCRCGGTQPCYKVIYFHDTSRRLNFEEAKEACRRDGGQLVSI ESEDEQKL 60
Db 22 ATGRLLSGQPVCRCGGTQPCYKVIYFHDTSRRLNFEEAKEACRRDGGQLVSI ESEDEQKL 81
QY 61 IEKFIENLLPSDGDWIGLRRREKQSNSTACQDLYAWTDGSIQFRNWWYVDEPSCGSEV 120
Db 82 IEKFIENLLPSDGDWIGLRRREKQSNSTACQDLYAWTDGSIQFRNWWYVDEPSCGSEV 141
QY 121 CVVMYHQPSAPAGIGGYPYMFQWDDRCNMKNFICKYSDEKPAVPSREAEGEETELTPV 180
Db 142 CVVMYHQPSAPAGIGGYPYMFQWDDRCNMKNFICKYSDEKPAVPSREAEGEETELTPV 201
QY 181 LPEETQEDAKKTFKESREAALNLAY 206
Db 202 LPEETQEDAKKTFKESREAALNLAY 227

RESULT 4
Q9Z209
ID Q9Z209 PRELIMINARY; PRT; 374 AA.
AC Q9Z209;
DT 01-MAY-1999 (TREMBlrel. 10, Created)
DT 01-MAY-1999 (TREMBlrel. 10, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE Layilin.
OS Cricetulus griseus (Chinese hamster).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
OC Cricetulus.
OX NCBI_TaxID=10029;
RN [1]
RP SEQUENCE FROM N.A.
RA Borowsky M.L., Hynes R.O.;
RT "Layilin, a novel talin-binding transmembrane protein homologous with C-type lectins, is localized in membrane ruffles.";
RL J. Cell Biol. 143:0-0(1998).
DR EMBL; AF093673; AAC68695.1; --
DR HSP; P06734; IHLI.
DR InterPro; IPR001304; Lectin_C.
DR Pfam; PF00059; lectin_c; 1.
DR SMART; SM00034; CLECT; 1.
DR PROSITE; PS50041; C_TYPE_LLECTIN_2; 1.
SQ SEQUENCE 374 AA; 42435 MW; 298A8BA24FB04E1C CRC64;

Query Match 84.8%; Score 945.5; DB 11; Length 374;
Best Local Similarity 85.9%; Pred. No. 5.5e-82;
Matches 176; Conservative 9; Mismatches 19; Indels 1; Gaps 1;

QY 3 GRLLSGQPVCRCGGTQPCYKVIYFHDTSRRLNFEEAKEACRRDGGQLVSI ESEDEQKLIE 62
Db 24 GRLLSGQLVCRGSTRPCYKVIYFHDAFQRLNFEEAKEACRRDGGQLVSI ETEDEQRLIE 83
QY 63 KFIENLLPSDGDWIGLRRREKQSNSTACQDLYAWTDGSIQFRNWWYVDEPSCGSEVCV 122
Db 84 KFIENLLASDGDWIGLRRLEVKQVNTACQDLYAWTDGSTSQFRNWWYVDEPSCGSEVCV 143
QY 123 VMYHQPSAPAGIGGYPYMFQWDDRCNMKNFICKYSDEKPA-VPSREAEGEETELTPVL 181

Db 144 VMYHQSAPPYGIGSYMFQWDDRCNMKNFICKYADEKPSITPSIRPGGEATEPPTPV 203
QY 182 PEETOEDAKTKFKESREAAALNAY 206
Db 204 PEETQEDTKETFKESREAAALNAY 228
RESULT 5
Q8VI31
ID Q8VI31 PRELIMINARY; PRT; 273 AA.
AC Q8VI31;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE C-type lectin protein MT75.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J;
RA Weng L., Smits P., Hubner R., Wouter's J., Merregaert J.;
RT "Mt75, a low expressed c-type lectin gene involving in
RT chondrogenesis";
RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF311699; AAL50354.1; -;
DR InterPro; IPR001304; Lectin_C.
DR Pfam; PF00059; lectin_c; 1.
DR SMART; SM00034; CLECT; 1.
DR FROSITE; PS50041; C_TYPE_LLECTIN_2; 1.
KW lectin.
SQ SEQUENCE 273 AA; 30303 MW; E052D933F244F4C7 CRC64;
Query Match 51.5%; Score 574; DB 11; Length 273;
Best Local Similarity 57.4%; Pred. No. 9.9e-47;
Matches 109; Conservative 26; Mismatches 41; Indels 14; Gaps 5;
QY 4 RLLSQPVCRGGTQPCYKVIYFHDTSRRLLNFEEAKEACRRDGGQLVSISEDEQKLIK 63
Db 23 RVVSQKVCFAADVKHPCYKMAFHELSRSVSFQEARLACESEGGVLLSLENAEQKLIES 82
QY 64 FIENLLP-----SDGDFWIGLRRREEKQSNSTACQDLYAWTDGSIQFRNMYVDEPSCGS 118
Db 83 MLQNLTKPGTGISDGFWIGLRLSGDGT-SGACPDLYQWSDGSSSQFRNMYTDEPSCGS 141
QY 119 EVCVVMYHQPSAPAGIGGPFQWDDRCNMKNFICKYSDE-KPAVPSRAEAGEETELT 177
Db 142 EKCVVMYHQPTANPGLGGPYLYQWDDRCNMKNHYICTYPEIHTPEPA-----EKPYLT 196
QY 178 TPVLPETQE 187
Db 197 NQ--PEETHE 204
RESULT 6
Q9W6E1
ID Q9W6E1 PRELIMINARY; PRT; 1290 AA.
AC Q9W6E1;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Neurocan core protein.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20309833; Pubmed=10851024;
RA Li H., Leung T.C., Hoffman S., Balsamo J., Lillian J.;

RT *Coordinate Regulation of Cadherin and Integrin Function by the
RT Chondroitin Sulfate Proteoglycan Neurocan.*;
RL J. Cell Biol. 149:1275-1288(2000).
DR EMBL; AF116856; AAD24546.2; -;
DR HSSP; P08709; IBF9.
DR InterPro; IPR000152; Asx_hydroxyl.
DR InterPro; IPR000561; EGF-like.
DR InterPro; IPR000742; EGF_2.
DR InterPro; IPR001881; EGF_Ca.
DR InterPro; IPR001438; EGF_II.
DR InterPro; IPR003599; Ig.
DR InterPro; IPR003006; Ig_MHC.
DR InterPro; IPR001304; Lectin_C.
DR InterPro; IPR000538; Link.
DR InterPro; IPR000436; Sushi_SCR_CCP.
DR Pfam; PF00008; EGF; 2.
DR Pfam; PF00047; Ig; 1.
DR Pfam; PF00059; lectin_c; 1.
DR Pfam; PF00084; sushi; 1.
DR Pfam; PF00193; xlink; 2.
DR PRINTS; PR00010; EGFBLLOOD.
DR PRODom; PD000918; Link; 2.
DR SMART; SM00032; CCP; 1.
DR SMART; SM00034; CLECT; 1.
DR SMART; SM00179; EGF_CA; 1.
DR SMART; SM00001; EGF_like; 1.
DR SMART; SM00409; IG; 1.
DR SMART; SM00445; LINK; 2.
DR PROSITE; PS00010; ASX_HYDROXYL; UNKNOWN_1.
DR PROSITE; PS00615; C_TYPE_LLECTIN_1; 1.
DR PROSITE; PS50041; C_TYPE_LLECTIN_2; 1.
DR PROSITE; PS00022; EGF_1; UNKNOWN_2.
DR PROSITE; PS01186; EGF_2; 1.
DR PROSITE; PS01187; EGF_CA; 1.
DR PROSITE; PS01241; LINK; 2.
KW Calcium-binding; EGF-like domain; Glycoprotein; Repeat.
SQ SEQUENCE 1290 AA; 138877 MW; 182BD86D0E40BE78 CRC64;
Query Match 16.6%; Score 185; DB 13; Length 1290;
Best Local Similarity 34.7%; Pred. No. 7.7e-09;
Matches 50; Conservative 18; Mismatches 42; Indels 34; Gaps 8;
QY 17 QRPCYKVIYFHDTSRRLLNFEEAKEACRRDGGQLVSISEDEQKLIKFIENLLPSDGF 76
Db 1064 QGHCYR--YF---SRRSWEDAERDCRRRAGHLTSIHSQEEHGFINSF-----GHENTW 1112
QY 77 IGLRRREEKQSNSTACQDLYAWTDGSIQFRNMYVDEPS---CGSEVCVVMY-HQPSAPA 132
Db 1113 IGLNDRIVEQD-----FQWTDNTGLQYENWRENQPDNFFAGGEDCVVLVSHE----- 1159
QY 133 GIGGPFQWDDRCNMKNFICK 156
Db 1160 -IG-----KWNDVPCNPNLPYICK 1177
RESULT 7
Q61830
ID Q61830 PRELIMINARY; PRT; 1456 AA.
AC Q61830;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Macrophage mannose receptor precursor.
GN MRC1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6;
RX MEDLINE=93043353; Pubmed=1421407;
RA Harris N., Rits M., Chang G., Ezekowitz R.B.;

RT "Characterization of the murine macrophage mannose receptor.";
RL Blood 80:2363-2373(1992).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6;
RA Super M.;
RL Submitted (APR-1992) to the EMBL/GenBank/DBJ databases.
DR EMBL; Z11974; CAA78028.1; -.
DR HSSP; P22897; LEGG.
DR MGD; MGI:97142; Mrcl.
DR InterPro; IPR000562; FN_Type_II.
DR InterPro; IPR001304; Lectin_C.
DR InterPro; IPR000772; Ricin_B_lectin.
DR Pfam; PF00040; fn2; 1.
DR Pfam; PF00059; lectin_c; 8.
DR Pfam; PF00652; Ricin_B_lectin; 2.
DR PRINTS; PR00013; FNTYPEII.
DR PRODOM; PD000995; FN_Type_II; 1.
DR SMART; SM00034; CLECT; 8.
DR SMART; SM00059; FN2; 1.
DR SMART; SM00458; RICIN; 1.
DR PROSITE; PS00615; C_TYPE_LLECTIN_1; 6.
DR PROSITE; PS50041; C_TYPE_LLECTIN_2; 8.
DR PROSITE; PS00023; FIBRONECTIN_2; 1.
DR PROSITE; PS50231; RICIN_B_LLECTIN; 1.
KW Receptor; Signal.
FT SIGNAL 1 18 POTENTIAL.
FT CHAIN 19 1456 MACROPHAGE MANNOSE RECEPTOR.
SQ SEQUENCE 1456 AA; 165065 MW; 4EBD3F1B8619A594 CRC64;

Query Match 16.0%; Score 178.5; DB 11; Length 1456;
Best Local Similarity 25.9%; Pred. No. 3.7e-08;
Matches 53; Conservative 33; Mismatches 66; Indels 53; Gaps 8;

QY 21 YKVIYFHDTSRRLLNFEEAKEACRRDGGQLVSIIESEDEQKLIKFIENLLPSDGFWIGLR 80
DB 807 YKDYQYYSKEKETMDNARRFCKKNFGDLATIKSEKKFLWKYI-NKNGGQSPYFIGML 865
QY 81 RREEKQSNSTACQDLYAWTDGSIQFRNWIYVDEPSCGS--EVCVVMYHQPSAPAGIGGPY 138
DB 866 ISMDKK-----FIWMDGSKVDFAWATGEPNFANDDENCVTMY-----TNSGF---- 908
QY 139 MFQWDDRCNMKNFICK---YSDEKPAVPSREAEGEETELTPVLPEETQE----- 187
DB 909 ---WNDINGYPNFIQORHNSINATAMP-----TTPTPGGCKREGWHLYKNK 954
QY 188 -----EDAKTKFKESREAAALNL 204
DB 955 CFKIFGFANEKKSWQDARQACKGL 979

Query Match 16.0%; Score 178.5; DB 11; Length 1456;
Best Local Similarity 25.9%; Pred. No. 3.7e-08;
Matches 53; Conservative 33; Mismatches 66; Indels 53; Gaps 8;

QY 21 YKVIYFHDTSRRLLNFEEAKEACRRDGGQLVSIIESEDEQKLIKFIENLLPSDGFWIGLR 80
DB 807 YKDYQYYSKEKETMDNARRFCKKNFGDLATIKSEKKFLWKYI-NKNGGQSPYFIGML 865
QY 81 RREEKQSNSTACQDLYAWTDGSIQFRNWIYVDEPSCGS--EVCVVMYHQPSAPAGIGGPY 138
DB 866 ISMDKK-----FIWMDGSKVDFAWATGEPNFANDDENCVTMY-----TNSGF---- 908
QY 139 MFQWDDRCNMKNFICK---YSDEKPAVPSREAEGEETELTPVLPEETQE----- 187
DB 909 ---WNDINGYPNFIQORHNSINATAMP-----TTPTPGGCKREGWHLYKNK 954
QY 188 -----EDAKTKFKESREAAALNL 204
DB 955 CFKIFGFANEKKSWQDARQACKGL 979

RESULT 8
Q25199
ID Q25199 PRELIMINARY; PRT; 1348 AA.
AC Q25199;
DT 01-NOV-1996 (TReMBLrel. 01, Created)
DT 01-MAY-1999 (TReMBLrel. 10, Last sequence update)
DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
DE Tyrosine kinase receptor.
OS Hydra attenuata (Hydra) (Hydra vulgaris).
OC Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydroida; Anthomedusae;
OC Hydridae; Hydra.
OX NCBI_TaxID=6087;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=IRVINE;
RX MEDLINE=20209407; PubMed=10744720;
RA Reidling J.C., Miller M.A., Steele R.E.;
RT "Sweet Tooth, a Novel Receptor Protein-tyrosine Kinase with C-type
Lectin-like Extracellular Domains.";
RL J. Biol. Chem. 275:10323-10330(2000).
DR EMBL; L22612; AAA29218.2; -.
DR HSSP; P11362; 1FGK.

DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR001304; Lectin_C.
DR InterPro; IPR001245; Tyr_pkinase.
DR Pfam; PF00059; lectin_c; 4.
DR Pfam; PF00069; pkinase; 1.
DR PRODOM; PD000001; Euk_pkinase; 2.
DR SMART; SM00034; CLECT; 4.
DR SMART; SM00219; TyrKC; 1.
DR PROSITE; PS00615; C_TYPE_LLECTIN_1; UNKNOWN_2.
DR PROSITE; PS50041; C_TYPE_LLECTIN_2; 4.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; UNKNOWN_1.
DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
KW ATP-binding; Kinase; Receptor; Transferase; Tyrosine-protein kinase.
SQ SEQUENCE 1348 AA; 156916 MW; 77D2122093227FEF CRC64;

Query Match 15.9%; Score 177.5; DB 5; Length 1348;
Best Local Similarity 28.0%; Pred. No. 4.2e-08;
Matches 47; Conservative 32; Mismatches 46; Indels 43; Gaps 8;

QY 6 LSGQPVC--RGGTQRPC-----YKVIYFHDTSRRLLNFEEAKEACRRDGGQLVSIIESEDE 57
DB 412 LSHRFICKVKRATNEYCAEGWTSYRIYCYFIYSIEFDWFKSFSSCQNIIGNLLSIENQEE 471
QY 58 QKLIKFIENLLPSDGD-FWIGLR-----REEKQSNSTACQDLYAWTDGSIQFRNWI 110
DB 472 ----NRFIENDLIKNDKYWIGLKNKIWDYLLKKNR-----FEWSDNTYTQFFNWI 518
QY 111 VDEP--SCGSEVCVVMYHQPSAPAGIGGPYMFQWDDRCNMKNFICK 156
DB 519 TNQPDNNGIESCVEMNYN-----GWSDKCKVLNGFICK 553

RESULT 9
Q9Y5P9
ID Q9Y5P9 PRELIMINARY; PRT; 1479 AA.
AC Q9Y5P9;
DT 01-NOV-1999 (TReMBLrel. 12, Created)
DT 01-NOV-1999 (TReMBLrel. 12, Last sequence update)
DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
DE Endocytic receptor Endo180.
GN ENDO180.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20148849; PubMed=10683150;
RA Sheikh H., Yarwood H., Ashworth A., Isacke C.;
RT "Endo180, an endocytic recycling glycoprotein related to the
macrophage mannose receptor is expressed on fibroblasts, endothelial
cells and macrophages and functions as a lectin receptor.";
RL J. Cell Sci. 113:1021-1032(2000).
DR EMBL; AF134838; AAD30280.1; -.
DR HSSP; P02751; 2FN2.
DR InterPro; IPR000562; FN_Type_II.
DR InterPro; IPR001304; Lectin_C.
DR InterPro; IPR000566; Lipocln_cytFABP.
DR InterPro; IPR000772; Ricin_B_lectin.
DR Pfam; PF00040; fn2; 1.
DR Pfam; PF00059; lectin_c; 8.
DR PRINTS; PR00013; FNTYPEII.
DR PRODOM; PD000995; FN_Type_II; 1.
DR SMART; SM00034; CLECT; 8.
DR SMART; SM00059; FN2; 1.
DR SMART; SM00458; RICIN; 1.
DR PROSITE; PS00615; C_TYPE_LLECTIN_1; 3.
DR PROSITE; PS50041; C_TYPE_LLECTIN_2; 8.
DR PROSITE; PS00023; FIBRONECTIN_2; UNKNOWN_1.
DR PROSITE; PS00213; LIPOCALIN; UNKNOWN_1.
DR PROSITE; PS50231; RICIN_B_LLECTIN; 1.
KW Receptor.

Db 90 QGHCYR--YF---AHRRAWEDAECDCRRRSGHLTSVHSPPEHSFINSF-----GHENTW 138

QY 77 IGLRRREKQSNSTACQDLYAWTDGSIQFRNMYVDEPS---CGSEVCVVMYHQPSAPAG 133

Db 139 IGLNDRIVERD-----FQWTDNTGLQFENWRENQPDNFFAGGEDCVVMVAHESG--- 187

QY 134 IGGPYMFQWDDRCNMKNFICK 156

Db 188 -----RWNDVPCNYNLPYVCK 203

RESULT 12

O14594

ID O14594 PRELIMINARY; PRT; 1321 AA.

AC O14594;

DT 01-JAN-1998 (TReMBLrel. 05, Created)

DT 01-MAY-1999 (TReMBLrel. 10, Last sequence update)

DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)

DE Neurocan (PGCN_HUMAN).

GN CSPG3 OR NEUR.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCBI_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=99013874; PubMed=9795216;

RA Prange C.K., Pennacchio L.A., Lieuallen K., Fan W., Lennon G.G.;

RT "Characterization of the human neurocan gene, CSPG3.";

RL Gene 221:199-205(1998).

RN [2]

RP SEQUENCE OF 1-990 FROM N.A.

RA Lamerdin J.E., McCreedy P.M., Adamson A.W., Burkhardt-Schultz K.,

RA Christensen M., Gordon L., Kyle A., Ramirez M., Stilwagen S.,

RA Garness J., Danganan L., Bruce R., Quan G., Montgomery M., Ow D.,

RA Kobayashi A., Olsen A.O., Carrano A.V.;

RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.

DR EMBL; AF026547; AAC80576.1; -.

DR EMBL; AC003110; AAB86655.1; -.

DR HSSP; P00740; 1EDM.

DR InterPro; IPR000152; Asx_hydroxyl.

DR InterPro; IPR000561; EGF-like.

DR InterPro; IPR000742; EGF_2.

DR InterPro; IPR001881; EGF_Ca.

DR InterPro; IPR003599; Ig.

DR InterPro; IPR003006; Ig_MHC.

DR InterPro; IPR001304; Lectin_C.

DR InterPro; IPR000538; Link.

DR InterPro; IPR000436; Sushi_SCR_CCP.

DR Pfam; PF00008; EGF; 2.

DR Pfam; PF00047; ig; 1.

DR Pfam; PF00059; lectin_c; 1.

DR Pfam; PF00084; sushi; 1.

DR Pfam; PF00193; Xlink; 2.

DR ProDom; PD000918; Link; 2.

DR SMART; SM00032; CCP; 1.

DR SMART; SM00034; CLECT; 1.

DR SMART; SM00179; EGF_CA; 1.

DR SMART; SM00001; EGF_like; 1.

DR SMART; SM00409; IG; 1.

DR SMART; SM00445; LINK; 2.

DR PROSITE; PS00010; ASX_HYDROXYL; UNKNOWN_1.

DR PROSITE; PS00615; C_TYPE_LLECTIN_1; 1.

DR PROSITE; PS50041; C_TYPE_LLECTIN_2; 1.

DR PROSITE; PS00022; EGF_1; UNKNOWN_3.

DR PROSITE; PS01186; EGF_2; 1.

DR PROSITE; PS01187; EGF_CA; 1.

DR PROSITE; PS01241; LINK; 2.

KW Calcium-binding; EGF-like domain; Glycoprotein; Repeat.

SQ SEQUENCE 1321 AA; 142972 MW; 2EF47F823DB980B8 CRC64;

Query Match 15.8%; Score 176; DB 4; Length 1321;

Best Local Similarity 31.5%; Pred. No. 5.7e-08;

Matches 45; Conservative 18; Mismatches 48; Indels 32; Gaps 6;

QY 17 QRPCYKVIYFHDTSRRLNFEAEKACRRDGGQLVSISEDEQKLIKFIENLLPSDGDWF 76

Db 1096 QGHCYR--YF---AHRRAWEDAECDCRRRSGHLTSVHSPPEHSFINSF-----GHENTW 1144

QY 77 IGLRRREKQSNSTACQDLYAWTDGSIQFRNMYVDEPS---CGSEVCVVMYHQPSAPAG 133

Db 1145 IGLNDRIVERD-----FQWTDNTGLQFENWRENQPDNFFAGGEDCVVMVAHESG--- 1193

QY 134 IGGPYMFQWDDRCNMKNFICK 156

Db 1194 -----RWNDVPCNYNLPYVCK 1209

RESULT 13

Q64449

ID Q64449 PRELIMINARY; PRT; 1479 AA.

AC Q64449;

DT 01-NOV-1996 (TReMBLrel. 01, Created)

DT 01-NOV-1996 (TReMBLrel. 01, Last sequence update)

DT 01-MAR-2002 (TReMBLrel. 20, Last annotation update)

DE Lectin lambda.

GN MRC2.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.

OX NCBI_TaxID=10090;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=96355501; Pubmed=8702911;

RA Wu K., Yuan J., Lasky L.A.;

RT "Characterization of a novel member of the macrophage mannose receptor

RT type C lectin family.";

RL J. Biol. Chem. 271:21323-21330(1996).

DR EMBL; U56734; AAC52729.1; -.

DR HSSP; P02751; 2FN2.

DR MGD; MGI:107818; Mrc2.

DR InterPro; IPR001128; Cytochrome_P450.

DR InterPro; IPR000562; FN_Type_II.

DR InterPro; IPR001304; Lectin_C.

DR InterPro; IPR000566; Lipocln_cytFABP.

DR InterPro; IPR000772; Ricin_B_lectin.

DR Pfam; PF00040; fn2; 1.

DR Pfam; PF00059; lectin_c; 8.

DR PRINTS; PR00013; FNTYPEII.

DR ProDom; PD000995; FN_Type_II; 1.

DR SMART; SM00034; CLECT; 8.

DR SMART; SM00059; FN2; 1.

DR SMART; SM00458; RICIN; 1.

DR PROSITE; PS00086; CYTOCHROME_P450; UNKNOWN_1.

DR PROSITE; PS00615; C_TYPE_LLECTIN_1; 3.

DR PROSITE; PS50041; C_TYPE_LLECTIN_2; 8.

DR PROSITE; PS00023; FIBRONECTIN_2; UNKNOWN_1.

DR PROSITE; PS00213; LIPOCALIN; UNKNOWN_1.

DR PROSITE; PS50231; RICIN_B_LLECTIN; 1.

SQ SEQUENCE 1479 AA; 167112 MW; 62D456E10B9B48C1 CRC64;

Query Match 15.3%; Score 170.5; DB 11; Length 1479;

Best Local Similarity 31.4%; Pred. No. 2.2e-07;

Matches 48; Conservative 22; Mismatches 48; Indels 35; Gaps 7;

QY 7 SQPVCVCGGTQRPCYKVIYFHDTSRRLNFEAEKACRRDGGQLVSISEDEQKLIKFIEN 66

Db 384 SWQPF-----QGHCYRL-----QAEKRSWQESKRACLRGGDLLSIHSMAELEFITKQIK 433

QY 67 NLLPSDGDWFIGLRRRREKQSNSTACQDLYAWTDGSIQFRNMYVDEPS---CGSEVCVW 123

Db 434 QEVE---ELWIGL-----NDLKQMNFEWSGSLVSTHWHFPFEPNFRDLSLEDCVT 482

QY 124 MYHQPSAPAGIGGPYMFQWDDRCNMKNFICK 156

Db 483 IW----GPEG-----RWNDSPCNQSLPSICK 504

